



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12N 15/53, 9/02, 5/10, 15/11, C07K 16/40, G01N 33/563, C12N 15/82, A01H 5/00</b>		<b>A1</b>	(11) International Publication Number: <b>WO 98/59057</b>
(21) International Application Number: <b>PCT/US98/13044</b>		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: <b>24 June 1998 (24.06.98)</b>		(43) International Publication Date: <b>30 December 1998 (30.12.98)</b>	
(30) Priority Data: <b>60/050,615 24 June 1997 (24.06.97) US</b>			
(71) Applicant: <b>THE GENERAL HOSPITAL CORPORATION [US/US]; Fruit Street, Boston, MA 02114 (US).</b>			
(71)(72) Applicants and Inventors: <b>GOODMAN, Howard, M. [US/US]; 10 The Ledges Road, Newton Centre, MA 02159 (US). NGUYEN, Long, V. [US/US]; 364 Rindge Avenue #4H, Cambridge, MA 02140 (US). CHIANG, Hui-Hwa [CN/US]; 34796 Dorado Common, Fremont, CA 94555 (US).</b>			
(74) Agents: <b>CIMBALA, Michele, A. et al.; Sterne, Kessler, Goldstein &amp; Fox P.L.L.C., Suite 600, 1100 New York Avenue, N.W., Washington, DC 20005-3934 (US).</b>			

## Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

With an indication in relation to deposited biological material furnished under Rule 13bis separately from the description.

## (54) Title: GA4 HOMOLOGUE DNA, PROTEIN AND METHODS OF USE

## (57) Abstract

The invention relates to GA4 homologue (GA4H) DNA and proteins encoded by *GA4H* DNA. GA4H is believed to be a member of the family of enzymes involved in the biosynthesis of the gibberellin family (GA) of plant growth hormones that promote various growth and developmental processes in higher plants, such as seed germination, stem elongation, flowering and fruiting. More specifically, the protein encoded by the *GA4H* loci may have similar function(s) to  $\beta$ -hydroxylases. The invention also relates to vectors containing the DNA and the expression of the protein encoded by the DNA of the invention in a host cell. Additional aspects of the invention are drawn to host cells transformed with the DNA or antisense sequence of the invention, the use of such host cells for the maintenance, or expression or inhibition of expression of the DNA of the invention and to transgenic plants containing DNA of the invention. Finally, the invention also relates to the use of the GA4 homologues to alter aspects of plant growth.

Ga4	W T T	G A L L H	T U R E	K M .	38
Ga4h2	S T T S	S T T L N	P K L .		35
Ga4h1	L A E T C .	.....	G L G T T	E F K L A	29
Ga4		P P P T G N	D H A	N O R	77
Ga4h2		A . . . . D T L	D I H	A V T	70
Ga4h1	A S R P S	V F D V E S	M N E	D S K	69
Ga4		L G L	G	S	117
Ga4h2		S R			110
Ga4h1		I S K	S S K T D M	S E E S	109
Ga4					156
Ga4h2					149
Ga4h1	D K E P	S P E	A D D	S Y R T N T	149
Ga4					196
Ga4h2					189
Ga4h1	H D T G I	R V D E E	R L Y C T	T V	189
Ga4	S . . L S	L M A L			234
Ga4h2	G . . P N	F Q G T			227
Ga4h1	H K L E K G	S K V G R G R			229
Ga4					274
Ga4h2					267
Ga4h1					269
Ga4					313
Ga4h2					306
Ga4h1					309
Ga4					353
Ga4h2					346
Ga4h1	A I S T G	A P S R I I	Q U Y G	V D M D A	349
Ga4					358
Ga4h2					347
Ga4h1					355

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## GA4 Homologue DNA, Protein and Methods of Use

5

### 10 *Field of the Invention*

The invention relates to the field of molecular biology and plant growth hormones, and especially to gibberellin synthesis.

### *Background of the Invention*

15 Gibberellins (GA) are a large family of tetracyclic triterpenoid plant growth hormones that promote various growth and developmental processes in higher plants. These processes include promotion of cell division and extension, seed germination, stem elongation, flowering and fruiting (Stowe, B.B. *et al.*, *Annu. Rev. Plant Physiol.* 8:181-216 (1957), Graebe, J.E. *Annu. Rev. Plant Physiol.* 38:419-465 (1987), Phillips *et al.*, *Plant. Physiol.* 108:1049-1057 (1995),  
20 Xu *et al.*, *Proc. Natl. Acad. Sci. USA* 92:6640-6444 (1995), Martin *et al.*, *Plant* 200:159-166 (1996)). Genes that can alter GA biosynthesis or sensitivity have had an impact on the development of new plant species and on agriculture in general.

25 A number of GA responsive dwarf mutants have been isolated from various plant species, such as maize, pea, and *Arabidopsis* (Phinney, B.O. *et al.*, "Chemical Genetics and the Gibberellin Pathway" in *Zea mays L. in Plant Growth Substance*, ed., P.F. Waering, New York: Academic (1982) pp. 101-110; Ingram, T.J. *et al.*, *Plant* 160:455-463 (1984); Koornneef, M., *Arabidopsis Inf.*

*Serv.* 15:17-20. (1978)). The dwarf mutants of maize (*dwarf-1*, *dwarf-2*, *dwarf-3*, *dwarf-5*) have been used to characterize the maize GA biosynthesis pathway by determining specific steps leading to biologically important metabolites (Phinney, B.O. *et al.*, "Chemical Genetics and the Gibberellin Pathway" in *Zea mays L. in Plant Growth Substance*, ed., P.F. Waering, New York: Academic (1982) pp. 101-110; Fujioka, S. *et al.*, *Plant Physiol.* 88:1367-1372 (1988)). Similar studies have been done with the dwarf mutants from a pea (*Pisum sativum L.*) (Ingram, T.J. *et al.*, *Plant* 160:455-463 (1984)). GA deficient mutants have also been isolated from *Arabidopsis* (*gal*, *ga2*, *ga3*, *ga4*, *ga5*) (Koornneef, M., *et al.*, *Theor. Appl. Genet.* 58:257-263 (1980)). The *Arabidopsis ga4* mutant, induced by ethyl methanesulfonate (EMS) mutagenesis, is a germinating, GA responsive, semidwarf plant whose phenotype can be restored to wild type by repeated application of exogenous GA (Koornneef, M. *et al.*, *Theor. Appl. Genet.* 58:257-263 (1980)).

The *GA4* gene encodes a  $\beta$ -hydroxylase in *Arabidopsis thaliana*. A mutant allele (*ga4*) blocks the conversion of 3- $\beta$ -hydroxy GAs, reducing the endogenous levels of GA<sub>1</sub>, GA<sub>8</sub> and GA<sub>4</sub> and increasing the endogenous levels of GA<sub>19</sub>, GA<sub>20</sub> and GA<sub>9</sub> (Talon, M. *et al.*, *Proc. Natl. Acad. Sci. USA* 87:7983-7987 (1990)). The reduced levels of the 3- $\beta$ -hydroxy GAs is the cause of the semidwarf phenotype of the *ga4* mutant. It has been suggested that the pea *le* mutant also encodes an altered form of 3- $\beta$ -hydroxylase (Ross, J.J. *et al.*, *Physiol. Plant.* 76:173-176 (1989)). The pea deactivation mutant, *sln*, causes an elongated slender phenotype (Ross *et al.*, *Plant J.* 7:512-523 (1995)). Thus,  $\beta$ -hydroxylase is clearly implicated in the process of plant growth.

Homologues of the *GA4* gene (*GA4H*) that encode GA4-homologue proteins (GA4H) are described in this application. Two specific homologues, GA4H1 and GA4H2 are exemplified. High levels of sequence homology between the *GA4H1*, *GA4H2* and *GA4* genes, as well as between the proteins encoded by these genes suggest that at least these two homologue proteins (GA4H1 and GA4H2) may have similar functions or catalyze similar reactions



in plants to that of GA4. Thus, the GA4H proteins should be useful for plant growth modulation.

### *Summary of the Invention*

5 The invention provides genes involved in gibberellin biosynthesis from which one can express and obtain proteins useful for the regulation of plant growth. Additionally, the invention provides for new DNA probes useful for obtaining additional GA4 homologue genes and proteins. Lastly, this invention provides methods of regulating plant growth.

10 The invention is first directed to GA4H DNA and proteins encoded by GA4H DNA.

The invention is further directed to GA4H antisense DNA, and to the GA4H antisense RNA transcribed from it.

15 The invention is further directed to vectors containing GA4H encoding DNA and to the expression of GA4H proteins encoded by GA4H DNA in a host cell.

The invention is further directed to vectors containing GA4H antisense DNA and to the expression of GA4H antisense RNA by the GA4H antisense DNA in a host cell.

20 The invention is further directed to host cells transformed with a GA4H encoding DNA of the invention, and to the use of such host cells for the maintenance of GA4H DNA or expression of a GA4H protein of the invention.

25 The invention is further directed to host cells transformed with a GA4H antisense DNA of the invention, and to the use of such host cells for the maintenance of the GA4H DNA or expression of the GA4H antisense RNA of the invention, as inhibitors of the expression of endogenous GA4H.

The invention is further directed to transgenic plants containing a GA4H-encoding or GA4H antisense DNA of the invention.

The invention is further directed to a method for altering plant growth, using a GA4H encoding or GA4H antisense DNA of the invention

The invention is further directed to a method for altering plant growth, using a recombinantly made GA4H protein of the invention.

5            Preferably, each of the above embodiments is directed to GA4H1 or GA4H2 or the cDNA or genomic DNA encoding the GA4 homologues, as well as the antisense DNA of *GA4H1* or *GA4H2*.

### ***Brief Description of the Drawings***

10            **Figure 1: *Sequence of the GA4 cDNA*** (Chiang, H.H., et al., *Plant Cell* 7:195-201 (1995)) (SEQ ID Nos. 1, 2, 3 and 4). The figure shows the locations from which DNA probes were generated. The underlined nucleotides (Unique probes) (SEQ ID No. 3) indicate the region specific to the *GA4* gene that was used as a probe. Probes (Homologous probes) (SEQ ID No. 4) generated from boxed nucleotides were used for isolation of the GA4 homologues.

15            **Figure 2A-2C: *DNA gel blots of Arabidopsis genomic DNA***. Figure 2A shows a blot that was hybridized to probes derived from the homologous region of the *GA4* gene (Figure 1) at low stringency (42°C). Figure 2B shows a blot that was hybridized at low stringency to probes derived from the unique region of the *GA4* gene (Figure 1). Figure 2C shows a blot that was hybridized at high  
20 stringency to probes derived from p3-1, GA4H1 gene (Figure 3), DNA. DNA in lanes 1, 2 and 3 was digested with *HindIII*, *BamHI*, and *EcoRI*, respectively. The predicted size (in kilobase pairs; kbp) of the three major hybridizing bands are shown on the left.

25            **Figure 3: *The restriction map of the genomic clone, pLVN103*** (ATCC accession no. 98436; Deposited at the American Type Culture Collection, 10801 University Boulevard, Manassas, VA 20110-2209, U.S.A.) under the terms of the

Budapest Treaty), containing two linked homologues of GA4. The plasmid pLVN103 contains the entire genomic insert from  $\lambda$ 3 but was cloned into pBSKS(+). Plasmid p3-1 is a subclone of  $\lambda$ 3 and carries the 2.1 kb *Hind*III fragment. This subclone contains most of the coding region of the GA4H1 gene.

5 The region containing both GA4H1 and GA4H2 genes are shown in more detail on the bottom of the figure. The arrows indicate the direction of transcription of these genes. The line indicates the noncoding area, and rectangular boxes represent the coding region of the DNA. Abbreviations: B, *Bam*HI; H, *Hind*III.

**Figure 4A-4B: Physical mapping the GA4H1 and GA4H2 genes by anchoring to mapped YACs.** PCR amplification of the GA4H1 (with GA-P2 and GA-P6 primers) (See Figure 6) and GA4H2 (with GA-P19 and GA-P20 primers) (See Figure 8) genes (For the primer sequences, see Example 1). Figure 4A shows an ethidium bromide stained gel of the PCR product. Figure 4B shows an Autoradiograph of a DNA blot of the gel in Figure 4A using probes derived from the genomic clone pLVN103. Primers GA-P19 and GA-P20 were used in lanes 1-2 and 4-6, while primers GA-P2 and GA-P6 were used in lanes 7-8 and 10-12. Molecular weight markers (1 kb DNA ladder) were loaded in lanes 3 and 9. DNA templates are: genomic clone pLVN103 (lanes 2 and 8); YAC CIC6C3 (lanes 1 and 7) of chromosome 2; CIC1E4 (lanes 4 and 10); CIC6C10 (lanes 5 and 11); and CIC10A11 (lanes 6 and 12).

**Figure 5: Nucleotide sequence (SEQ ID No. 5) of the GA4H1 RT-PCR product (cDNA).** The predicted start (ATG) and stop (taa) codons are present at nucleotide nos. 44 and 1109, respectively. The intron is located at nucleotide no. 513 and is represented by a filled triangle (▼). Underlined nucleotides indicate the start (ATG) and stop (taa) codons. Lower case nucleotides represent 5' and 3' untranslated regions. A "G" at nucleotide no. 1059, indicated with an asterisk (\*), does not agree with the genomic DNA at this position. The number on the left indicates the nucleotide position.

**Figure 6: The genomic sequence of the GA4H1 gene** (SEQ ID No. 6).

Upper and lower case letters represent the coding and noncoding regions of the gene, respectively. The predicted translated protein sequence (SEQ ID No. 7) is shown below its corresponding nucleotide sequence. Arrows represent primers used in either PCR or RT-PCR analyses. The nucleotide and the amino acid positions are shown on the right.

**Figure 7: Nucleotide sequence of the GA4H2 RT-PCR product (cDNA)**

(SEQ ID No. 8). The predicted start (ATG) and stop (taa) codons are present at sequence nos. 49 and 1190, respectively. The intron is located at sequence no. 518. The number on the left indicates the nucleotide position.

**Figure 8: Genomic sequence of the GA4H2 gene** (SEQ ID No. 9).

Upper and lower case letters represent the coding and noncoding regions of the gene, respectively. The predicted translated protein sequence (SEQ ID No. 10) is shown below its corresponding nucleotide sequence. Arrows represent primers used in either PCR or RT-PCR analyses. The position of the nucleotide and the amino acid are shown on the right.

**Figure 9:** Alignment of GA4, GA4H1 and GA4H2 proteins. Both Pileup and Prettybox (Genetics Computer Group, Wisconsin, MA, U.S.A.) commands were used to generate this alignment. The position of the amino acid is shown on the right.

**Figure 10:** Amino acid sequence identity and similarity between GA4 (SEQ ID No. 2), GA4H1 (SEQ ID No. 7), GA4H2 (SEQ ID No. 10) and some other related 2-oxoacid-dependent dioxygenases (2-ODD). The percentage of sequence identity and similarity (in parenthesis) were generated using the GAP software of the GCG package. Shaded boxes indicate the putative GA4 gene family in *Arabidopsis*. Abbreviations: GA5, *Arabidopsis* GA<sub>20</sub>-oxidase (accession

number X83379); F3H, *Zea mays* flavanone-3- $\beta$ -hydroxylase (accession number U04434); FLS, potato flavanol synthase (accession number X92178); ANS, apple anthocyanidin hydroxylase (accession number S33144); EFE, tobacco ethylene forming enzyme (accession number Z29529). Accession number refer to GENBANK.

**Figure 11A-11B:** GA4H1 gene expressed in the flowers and shoot meristems. One-tenth of the PCR product of each sample was electrophoresed on an agarose gel and then stained with ethidium bromide (Figure 11A). A DNA blot of the gel in Figure 11A was probed with GA4H1 specific DNA (Figure 11B). Primers, GA-P13 and GA-P17, were used to amplify the 220 bp cDNA and 630 bp genomic DNA of the GA4H1 gene. Primers Tua4F/ Tua4R were used as an internal control that amplified the 320 bp cDNA of the  $\alpha$ -tubulin 4 gene (*TUA4*). DNA templates of pLVN115 (lane 1), pCD7 (lane 2), and pLVN103 (lane 3) were used in the PCR amplification. First strand cDNA templates of floral shoots (lane 5), leaves (lane 6), roots (lane 7), and siliques (lane 8) were subjected to RT-PCR. The 123 bp BRL DNA marker is present in lane 4.

**Figure 12A-12B:** GA4H2 gene expressed predominantly in the roots. One-tenth of the PCR product from each sample was separated on agarose gel and then stained with ethidium bromide (Figure 12A). The DNA gel blot shown in Figure 12A was probed with the GA4H2 specific probes (Figure 12B). Primers, GA-P18 and GA-P20, were used to amplify the 440 bp cDNA and 860 bp genomic DNA of the GA4H2 gene. The same primer pair of the *TUA4* gene was also used as an internal control during the RT-PCR. RNA templates of siliques (lane 1), roots (lane 2), leaves (lane 3), and floral shoots (lane 4) were subjected to RT-PCR. DNA templates of pLVN103 (lane 6), pCD7 (lane 7), and pLVN107 (lane 8) were used in the PCR amplification. The 123 bp BRL DNA marker is present in lane 5.

**Figure 13.** Phenotype of transgenic plants expressing the sense and antisense of the *GA4H1* gene.

### *Definitions*

5 "GA<sub>n</sub>" (with a number subscripted), refers to the "gibberellin A<sub>n</sub>" compound. The chemical structures of some of the gibberellin A<sub>n</sub>'s are presented in Moritz, T. *et al.*, *Plant* 193:1-8 (1994). GA without a subscript, e.g. GA1 refers to enzymes presumably involved in the gibberellin biosynthetic pathway.

Italicized, uppercase names, such as "*GA4* or *GA4H*," refer to the wild type gene. Italicized, lowercase names such as "*ga4*" refer to the mutant gene.

10 Uppercase names, such as "*GA4H*," refer to the protein, DNA or RNA encoded by a *GA4H* gene, while lowercase names, such as "*ga4*," refer to the protein, DNA or RNA encoded by a mutant, such as the mutant *ga4* gene.

GA4H refers to any GA4 homologue, while GA4H1 and GA4H2 refers to the homologues of GA4 shown in figures 6 and 8, or minor variations of these homologues or their cDNAs (Figures 5 and 7) . Such minor variations may include, but are not limited to substitution of conservative amino acids or degenerate substitutions in the DNA encoding the amino acid sequence of GA4H1 and GA4H2. Such variation may also be referred to as "substantially similar" molecules.

20 A *unique* probe should be understood to be a probe that contains a DNA sequence unique to GA4 DNA and that can be used to pull out the GA4 DNA. A "unique" probe sequence is indicated in Figure 1 by underlining. A *homologue* probe contains a DNA sequence homologous to a sequence found in GA4 homologue DNA. A "homologous" probe sequence is indicated in Figure 1 by the boxed nucleotide sequence and can be used to obtain GA4H DNA.

25 *Plant* should be understood as referring to a multicellular differentiated organism capable of photosynthesis including angiosperms (monocots and dicots) and gymnosperms.

*Plant cell* should be understood as referring to the structural and physiological unit of plants. The term "plant cell" refers to any cell which is either part of or derived from a plant. Some examples of cells encompassed by the present invention include differentiated cells that are part of a living plant; differentiated cells in culture; undifferentiated cells in culture; the cells of undifferentiated tissue such as callus or tumors.

*Plant cell progeny* should be understood as referring to any cell or tissue derived from plant cells including callus; plant parts such as stems, roots, fruits, leaves or flowers; plants; plant seed; pollen; and plant embryos.

*Propagules* should be understood as referring to any plant material capable of being sexually or asexually propagated, or being propagated *in vivo* or *in vitro*. Such propagules preferably consist of the protoplasts, cells, calli, tissues, embryos or seeds of the regenerated plants.

*Transgenic plant* should be understood as referring to a plant having stably incorporated exogenous DNA (i.e. DNA not normally found) in its genetic material. The term also includes exogenous DNA which may be introduced into a cell or protoplast in various forms, including, for example, naked DNA in circular, linear or supercoiled form, DNA contained in nucleosomes or chromosomes or nuclei or parts thereof, DNA complexed or associated with other molecules, DNA enclosed in liposomes, spheroplasts, cells or protoplasts.

*Purified* as it refers to preparations made from biological cells or hosts should be understood to mean any cell extract containing the indicated DNA or protein including a crude extract of the DNA or protein of interest. For example, in the case of a protein, a purified preparation can be obtained following an individual technique or a series of preparative or biochemical techniques and the DNA or protein of interest can be present at various degrees of purity in these preparations. The procedures may include for example, but are not limited to, ammonium sulfate fractionation, gel filtration, ion exchange chromatography, affinity chromatography, density gradient centrifugation and electrophoresis.

A preparation of DNA or protein that is "pure" or "isolated" should be understood to mean a preparation free from naturally occurring materials with which such DNA or protein is normally associated in nature. "Essentially pure" should be understood to mean a "highly" purified preparation that contains at least 95% of the DNA or protein of interest.

A *cell extract* that contains the DNA or protein of interest should be understood to mean a homogenate preparation or cell-free preparation obtained from cells that express the protein or contain the DNA of interest. The term "cell extract" is intended to include culture media, especially spent culture media from which the cells have been removed.

A *fragment* of a molecule should be understood as referring to a shortened sequence of an amino acid or nucleotide sequence that retains one or more desired chemical or biological properties of the full-length sequence such that use of the full-length sequence.

A *functional derivative* of GA4H (or GA4) should be understood as referring to a protein, or DNA encoding a protein, that possesses a biological activity that is substantially similar to the biological activity of GA4H (or GA4). A functional derivative may or may not contain post-translational modifications such as covalently linked carbohydrate, depending on the necessity of such modifications for the performance of a specific function. The term "functional derivative" is intended to include the "fragments," "variants," "analogues," or "chemical derivatives" of a molecule. The derivative retains at least one of the naturally-occurring functions of the parent gene or protein. The function can be any of the regulatory gene functions or any of the function(s) of the finally processed protein. The degree of activity of the function need not be quantitatively identical as long as the qualitative function is substantially similar.

A *mutation* should be understood as referring to a detectable change in the genetic material which may be transmitted to daughter cells and possibly even to succeeding generations giving rise to mutant cells or mutant organisms. If the



descendants of a mutant cell give rise only to somatic cells in multicellular organisms, a mutant spot or area of cells arises. Mutations in the germ line of sexually reproducing organisms may be transmitted by the gametes to the next generation resulting in an individual with the new mutant condition in both its somatic and germ cells. A mutation may be any (or a combination of) detectable, unnatural change affecting the chemical or physical constitution, mutability, replication, phenotypic function, or recombination of one or more deoxyribonucleotides; nucleotides may be added, deleted, substituted for, inverted, or transposed to new positions with and without inversion. Mutations may occur spontaneously and can be induced experimentally by application of mutagens. A mutant variation of a nucleic acid molecule results from a mutation. A mutant polypeptide may result from a mutant nucleic acid molecule.

A *species* should be understood as referring to a group of actually or potentially interbreeding natural populations. A species variation within a nucleic acid molecule or protein is a change in the nucleic acid or amino acid sequence that occurs among species and may be determined by DNA sequencing of the molecule in question.

A preparation that is *substantially free of other A. thaliana DNA (or protein)* should be understood as referring to a preparation wherein the only *A. thaliana* DNA (or protein) is that of the recited *A. thaliana* DNA (or protein). Though proteins may be present in the sample which are homologous to other *A. thaliana* proteins, the sample is still said to be substantially free of such other *A. thaliana* DNA (or protein) as long as the homologous proteins contained in the sample are not expressed from genes obtained from *A. thaliana*.

A *DNA construct* should be understood as referring to a recombinant, man-made DNA, linear or circular.

*T-DNA* (transferred DNA) should be understood as referring to a segment or fragment of Ti (tumor-inducing) plasmid DNA which integrates into the plant nuclear DNA.

*Stringent hybridization* conditions should be understood to be those conditions normally used by one of skill in the art to establish at least a 90% homology between complementary pieces of DNA or DNA and RNA. Lesser homologies, such as at least 70% homology or preferably at least 80% may also be desired and obtained by varying the hybridization conditions.

There are only three requirements for hybridization to a denatured strand of DNA to occur. (1) There must be complementary single strands in the sample. (2) The ionic strength of the solution of single-stranded DNA must be fairly high so that the bases can approach one another; operationally, this means greater than 0.2M. (3) The DNA concentration must be high enough for intermolecular collisions to occur at a reasonable frequency. The third condition only affects the rate, not whether renaturation/hybridization will occur.

Conditions routinely used by those of skill in the art are set out in readily available procedure texts, *e.g.*, *Current Protocols in Molecular Biology*, Vol. I, Chap. 2.10, John Wiley & Sons, Publishers (1994) or Sambrook *et al.*, *Molecular Cloning*, Cold Spring Harbor (1989), incorporated herein by reference. As would be known by one of skill in the art, the ultimate hybridization stringency reflects both the actual hybridization conditions as well as the washing conditions following the hybridization, and one of skill in the art would know the appropriate manner in which to change these conditions to obtain a desired result.

For example, a prehybridization solution should contain sufficient salt and nonspecific DNA to allow for hybridization to non-specific sites on the solid matrix, at the desired temperature and in the desired prehybridization time. For example, for stringent hybridization, such prehybridization solution could contain 6x sodium chloride/sodium citrate (1xSSC is 0.15 M NaCl, 0.015 M Na citrate; pH 7.0), 5x Denhardt's solution, 0.05% sodium pyrophosphate and 100 µg per ml of herring sperm DNA. An appropriate stringent hybridization mixture might then contain 6x SSC, 1x Denhardt's solution, 100 µg per ml of yeast tRNA and 0.05% sodium pyrophosphate.

Alternative conditions for DNA-DNA analysis could entail the following:

- 1) prehybridization at room temperature and hybridization at 68°C;
- 2) washing with 0.2x SSC/0.1% SDS at room temperature;
- 3) as desired, additional washes at 0.2x SSC/0.1% SDS at 42°C (moderate-stringency wash); or
- 5 4) as desired, additional washes at 0.1x SSC/0.1% SDS at 68°C (high stringency).

Known hybridization mixtures, *e.g.*, that of Church and Gilbert, *Proc. Natl. Acad. Sci. USA* 81:1991-1995 (1984), comprising the following composition may also be used: 1% crystalline grade bovine serum albumin/1mM EDTA/0.5M NaHPO<sub>4</sub>, pH 7.2/7% SDS. Additional, alternative but similar  
10 reaction conditions can also be found in Sambrook *et al.*, *Molecular Cloning*, Cold Spring Harbor (1989). Formamide may also be included in prehybridization/hybridization solutions as desired.

It should be understood that these conditions are not meant to be definitive  
15 or limiting and may be adjusted as required by those of ordinary skill in the art to accomplish the desired objective.

A *vector* should be understood to be a DNA element used as a vehicle for cloning or expressing a desired sequence, such as a gene of the invention, in a host.

20 A *host* or *host cell* should be understood to be a cell in which a recombinant sequence, such as a sequence encoding a GA4H DNA of the invention, is incorporated and expressed. A *GA4H* gene of the invention or the antisense of the gene may be introduced into a host cell as part of a vector by transformation. Both the sense and the antisense DNA sequences are present in  
25 the same host cell since DNA is double stranded. The direction of transcription, however, as directed by an operably linked promoter as designed by the artisan, dictates which of the two strands is ultimately transcribed into RNA.

### ***Detailed Description***

30 The process for genetically engineering GA4H protein sequences,

according to the invention, is facilitated through the cloning of genetic sequences that are capable of encoding GA4H proteins and through the expression of such genetic sequences. As used herein, the term "genetic sequence" is intended to refer to a nucleic acid sequence (preferably DNA). Genetic sequences that are capable of encoding GA4H proteins can be derived from a variety of sources. These sources include genomic DNA, RNA, cDNA, synthetic DNA, and combinations thereof. The preferred source of the GA4H genomic DNA is a plant genomic library and most preferably an *Arabidopsis* genomic library. A more preferred source of the GA4H cDNA is a plant cDNA library and most preferably an *Arabidopsis* cDNA library made from silique mRNA, although the message is ubiquitously expressed in the root, leaf and flower of plants. This invention, however, is not meant to be limited to GA4H homologues from only the plant genus *Arabidopsis*.

Methods for obtaining and screening genomic libraries are well known in the art. An example of obtaining and screening a genomic library which is not meant to be limiting follows. Additional methods may be found in Example 1 of the specification.

One may begin with a CsCl DNA preparation and partially digests it with Sau3AI. After digestion, a partial fill-in reaction is performed. The reaction mixture for the partial fill-in is as follows:

40 µl	DNA
6 µl	Sau3AI buffer (10X)
2.5 µl	0.1 M DTT
1 µl	100 mM dATP
1 µl	100 mM dGTP
5 µl	Klenow enzyme
4.5 µl	H <sub>2</sub> O

After 30 minutes at 37°C the reaction is terminated with phenol-chloroform and the DNA is obtained. The DNA is then loaded on a 0.7% low melting point agarose gel and after electrophoresing, bands between 10 and 23 kb are cut out

-15-

from the gel. The gel with the cut-out bands is then melted at 67°C. The isolated DNA is then placed in the following ligation mixture:

5                   2 µl     Lambda Fix II, pre-digested arms (2 µg)  
                  1 µg     genomic DNA, partial fill-in  
                  0.5 µl    10x ligation buffer  
                  0.5 µl    10 mM ATP (pH 7.05)  
                  0.5 µl    T4 DNA ligase  
                  ~1.5 µl H<sub>2</sub>O (to 5µl final volume)

10                   Following ligation overnight at 4°C, the DNA is packaged using GIGAPACK II GOLD.

                  Plaque lifts are made using Hybond filters (Amersham Corp.), which were then autoclaved for 2 min. Filters were hybridized with probes as described for DNA and RNA gel blot analysis below.

15                   Bacteriophage λ DNA is prepared from ER1458 lysates according to the mini-prep method of Grossberger, D., Nucl. Acids. Res. 15:6737 (1987). DNA fragments are subcloned into pBluescript KS<sup>-</sup> vectors (Stratagene) and used to transform JM109.

20                   Double stranded DNA is isolated from plasmid clones and purified by CsCl banding. Sequencing is performed using the ABI PRISM dye terminator cycle sequencing kit and the products are separated and detected on the ABI 377 (Perkin Elmer). Sequence analysis is performed using the Sequence Analysis Software package (Genetics Computer Group, Inc., Madison, WI) and the Blast network service of the National Center for Biotechnology Information (Bethesda, MD).

25                   Electrophoresis of DNA is in Tris-Acetate-EDTA buffer with subsequent transfer in 25 mM NaHPO<sub>4</sub> to Biotrans filters (International Chemical and Nuclear Corp.). Electrophoresis of RNA samples is in agarose gels containing RNAase inhibitor using MOPS/EDTA buffer and transferred to filters as for DNA. Filters were UV-crosslinked using a Stratalinker (Stratagene) and baked  
30                   for 1 hr at 80°C.

Radioactive probes are separated from unincorporated nucleotides using a 1-ml Sephadex G-50 spin column and denatured in a microwave oven (Stroop, W.G. et al., Anal. Biochem. 182:222-225 (1989)). Prehybridization for 1 hr and hybridization overnight is performed at 65°C in the hybridization buffer described by Church, G.M. et al., Proc. Natl. Acad. Sci. USA 81:1991-1995 (1984)). Filters are washed once for 15 min in 2xSSC at room temperature, then two times for 30 min in 0.1xSSC and 0.1%SDS at 60°C. The damp filters are autoradiographed at -80°C using intensifying screens. Filters are stripped twice in 2mM Tris-HCl, pH8.0, 1mM EDTA, 0.2% SDS at 70°C for 30 min prior to reprobing (Church, G.M. et al., Proc. Natl. Acad. Sci. USA 81:1991-1995 (1984)).

The recombinant GA4H cDNA of the invention will not include naturally occurring introns if the cDNA is made using mature GA4H mRNA as a template. Genomic DNA may or may not include naturally occurring introns. Moreover, such genomic DNA may be obtained in association with the homologous (isolated from the same source; native) 5' promoter region of the *GA4H* gene sequences and/or with the homologous 3' transcriptional termination region. Further, such genomic DNA may be obtained in association with the genetic sequences that provide the homologous 5' non-translated region of the GA4 mRNA and/or with the genetic sequences which provide the homologous 3' non-translated region.

Due to the degeneracy of nucleotide coding sequences, and to the fact that the DNA code is known, all other DNA sequences which encode the same amino acid sequence as depicted for example, in Figure 6 [SEQ ID No. 7] can be determined and used in the practice of the present invention. Additionally, those sequences that hybridize to for example, to a GA4H sequence such as SEQ. ID Nos. 5 or 6, under stringent conditions are also useful in the practice of the present invention.

A DNA sequence encoding GA4H protein or GA4H antisense RNA can be inserted into a DNA vector in accordance with conventional techniques,

including blunt-ending or staggered-ending termini for ligation, restriction enzyme digestion to provide appropriate termini, filling in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and ligation with appropriate ligases. In one embodiment of the invention, expression  
5 vectors are provided that are capable of expressing GA4H mRNA or antisense RNA. Vectors for propagating a given sequence in a variety of host systems are well known and can readily be altered by one of skill in the art such that the vector will contain DNA or RNA encoding the desired genetic sequence and will be propagated in a desired host. Such vectors include plasmids and viruses and  
10 such hosts include eukaryotic organisms and cells, for example plant, yeast, insect, plant, mouse or human cells, and prokaryotic organisms, for example *E. coli* and *B. subtilis*. Shuttle vectors in which the desired genetic sequence is "maintained" in an available form before being extracted and transformed into a second host for expression are also useful DNA constructs envisioned as carrying  
15 the DNA of the invention.

A nucleic acid molecule, such as DNA, is said to be "capable of expressing" a polypeptide or antisense sequence if it contains a nucleotide sequence that encodes such polypeptide or antisense sequence and transcriptional and, if necessary, translational regulatory information operably linked to the  
20 nucleotide sequences that encode the polypeptide or antisense sequence.

Two DNA sequences (such as a promoter region sequence and the *GA4H* gene encoding or antisense sequence) are said to be *operably linked* if the nature of the linkage between the two DNA sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the  
25 promoter region sequence to direct the transcription of the desired sequence, or (3) interfere with the ability of the desired sequence to be transcribed by the promoter region sequence. Thus, a promoter region would be operably linked to a desired DNA sequence if the promoter were capable of effecting transcription of that DNA sequence.

In one embodiment of the invention, a vector is employed that is capable of integrating the desired gene sequences into the host cell chromosome. Cells that have stably integrated the introduced DNA into their chromosomes can be selected by also introducing one or more markers which allow for selection of host cells which contain the expression vector. The marker may provide for prototrophy to an auxotrophic host, biocide resistance, e.g., antibiotics, or heavy metals, such as copper, or the like. The selectable marker gene sequence can either be directly linked to the DNA gene sequences to be expressed, or introduced into the same cell by co-transfection.

In another embodiment, the introduced sequence will be incorporated into a plasmid or viral vector capable of autonomous replication in the recipient host. Any of a wide variety of vectors may be employed for this purpose. Factors of importance in selecting a particular plasmid or viral vector include: the ease with which recipient cells that contain the vector may be recognized and selected from those recipient cells which do not contain the vector; the number of copies of the vector which are desired in a particular host; and whether it is desirable to be able to "shuttle" the vector between host cells of different species.

The present invention also encompasses the expression of the GA4H protein (or a functional derivative thereof) in either prokaryotic or eukaryotic cells. Preferred prokaryotic hosts include bacteria such as *E. coli*, *Bacillus*, *Streptomyces*, *Pseudomonas*, *Salmonella*, *Serratia*, etc. The most preferred prokaryotic host is *E. coli*. Bacterial hosts of particular interest include *E. coli* K12 strain 294 (ATCC 31446), *E. coli*  $\chi$ 1776 (ATCC 31537), *E. coli* W3110 (F<sup>-</sup>, lambda<sup>-</sup>, prototrophic (ATCC 27325)), and other enterobacterium such as *Salmonella typhimurium* or *Serratia marcescens*, and various *Pseudomonas* species. Under such conditions, the GA4H gene product will not be glycosylated. The prokaryotic host must be compatible with the replicon and control sequences in the expression plasmid.

Hosts can be utilized for production of the desired genetic sequence, or GA4H protein, using conventional methods, such as by growth in shake flasks,



fermentors, tissue culture plates or bottles. Alternatively, multicellular organisms such as a plant might be used.

DNA encoding the desired protein is preferably operably linked to a promoter region, a transcription initiation site, and a transcription termination sequence, functional in plants. Any of a number of promoters which direct transcription in a plant cell is suitable. The promoter can be either constitutive or inducible. Some examples of promoters functional in plants include the nopaline synthase promoter and other promoters derived from native Ti plasmids, viral promoters including the 35S and 19S RNA promoters of cauliflower mosaic virus (Odell *et al.*, *Nature* 313:810-812 (1985)), and numerous plant promoters.

Alternative promoters that may be used include nos, ocs, and CaMV promoters. Overproducing plant promoters may also be used. Such promoters, operably linked to the *GA4H* gene, should increase the expression of the GA4 protein. Overproducing plant promoters that may be used in this invention include the promoter of the small subunit (ss) of ribulose-1,5-biphosphate carboxylase from soybean (Berry-Lowe *et al.*, *J. Molecular and App. Gen.* 1:483-498 (1982), and the promoter of the chlorophyll a/b binding protein. These two promoters are known to be light-induced in eukaryotic plant cells (see, for example, *Genetic Engineering of Plants, an Agricultural Perspective*, A. Cashmore, Plenum, New York 1983, pages 29-38; Corruzi, G. *et al.*, *J. of Biol. Chem.* 258:1399 (1983); and Dunsmuir, P. *et al.*, *J. of Mol. and Applied Genet.* 2:285 (1983)).

To express the *GA4H* gene (or a functional derivative thereof) in a prokaryotic cell (such as, for example, *E. coli*, *B. subtilis*, *Pseudomonas*, *Streptomyces*, etc.), it is necessary to operably link the *GA4H* gene encoding sequence to a functional prokaryotic promoter. Such promoters may be either constitutive or, more preferably, regulatable (i.e., inducible or derepressible). Examples of constitutive promoters include the *int* promoter of bacteriophage  $\lambda$ , the *bla* promoter of the  $\beta$ -lactamase gene sequence of pBR322, and the CAT promoter of the chloramphenicol acetyl transferase gene sequence of pBR325,

etc. Examples of inducible prokaryotic promoters include the major right and left promoters of bacteriophage  $\lambda$  ( $P_L$  and  $P_R$ ), the *trp*, *recA*, *lacZ*, *lacI*, and *gal* promoters of *E. coli*, the  $\alpha$ -amylase (Ulmanen, I., *et al.*, *J. Bacteriol.* 162:176-182 (1985)) and the  $\zeta$ -28-specific promoters of *B. subtilis* (Gilman, M.Z., *et al.*, *Gene sequence* 32:11-20 (1984)), the promoters of the bacteriophages of *Bacillus* (Gryczan, T.J., In: *The Molecular Biology of the Bacilli*, Academic Press, Inc., NY (1982)), and *Streptomyces* promoters (Ward, J.M., *et al.*, *Mol. Gen. Genet.* 203:468-478 (1986)).

Prokaryotic promoters are reviewed by Glick, B.R., (*J. Ind. Microbiol.* 1:277-282 (1987)); Cenatiempo, Y. (*Biochimie* 68:505-516 (1986)); and Gottesman, S. (*Ann. Rev. Genet.* 18:415-442 (1984)).

Proper expression in a prokaryotic cell also requires the presence of a ribosome binding site upstream of the gene sequence-encoding sequence. Such ribosome binding sites are disclosed, for example, by Gold, L., *et al.* (*Ann. Rev. Microbiol.* 35:365-404 (1981)).

Preferred eukaryotic hosts include yeast, fungi, insect cells, mammalian cells either *in vivo*, or in tissue culture. Mammalian cells that can be useful as hosts include cells of fibroblast origin such as VERO or CHO-K1, or cells of lymphoid origin, such as the hybridoma SP2/O-AG14 or the myeloma P3x63Sg8, and their derivatives. Preferred mammalian host cells include SP2/0 and J558L, as well as neuroblastoma cell lines such as IMR 332 that may provide better capacities for correct post-translational processing.

For a mammalian host, several possible vector systems are available for the expression of the *GA4H* gene. A wide variety of transcriptional and translational regulatory sequences may be employed, depending upon the nature of the host. The transcriptional and translational regulatory signals may be derived from viral sources, such as adenovirus, bovine papilloma virus, Simian virus, or the like, where the regulatory signals are associated with a particular gene sequence which has a high level of expression. Alternatively, promoters from mammalian expression products, such as actin, collagen, myosin, etc., may

be employed. Transcriptional initiation regulatory signals may be selected that allow for repression or activation, so that expression of the gene sequences can be modulated. Of interest are regulatory signals which are temperature-sensitive so that by varying the temperature, expression can be repressed or initiated, or are  
5 subject to chemical (such as metabolite) regulation.

Yeast provides substantial advantages in that it can also carry out post-translational peptide modifications. A number of recombinant DNA strategies exist that utilize strong promoter sequences and high copy number plasmids that can be utilized for production of the desired proteins in yeast. Yeast recognizes  
10 leader sequences on cloned mammalian gene sequence products and secretes peptides bearing leader sequences (i.e., pre-peptides).

Any of a series of yeast gene sequence expression systems incorporating promoter and termination elements from the actively expressed gene sequences coding for glycolytic enzymes produced in large quantities when yeast are grown  
15 in medium rich in glucose can be utilized. Known glycolytic gene sequences can also provide very efficient transcriptional control signals. For example, the promoter and terminator signals of the phosphoglycerate kinase gene sequence can be utilized.

Another preferred host is insect cells, for example the *Drosophila* larvae. Using insect cells as hosts, the *Drosophila* alcohol dehydrogenase promoter can  
20 be used (Rubin, G.M., *Science* 240:1453-1459 (1988)). Alternatively, baculovirus vectors can be engineered to express large amounts of the *GAl* gene in insects cells (Jasny, B.R., *Science* 238:1653 (1987); Miller, D.W., *et al.*, in *Genetic Engineering* (1986), Setlow, J.K., *et al.*, eds., *Plenum*, Vol. 8, pp. 277-  
25 297).

As discussed above, expression of the *GA4H* gene in eukaryotic hosts requires the use of eukaryotic regulatory regions. Such regions will, in general, include a promoter region sufficient to direct the initiation of RNA synthesis. Preferred eukaryotic promoters include the promoter of the mouse  
30 metallothionine I gene sequence (Hamer, D., *et al.*, *J. Mol. Appl. Gen.* 1:273-288

(1982)); the TK promoter of Herpes virus (McKnight, S., *Cell* 31:355-365 (1982)); the SV40 early promoter (Benoist, C., *et al.*, *Nature (London)* 290:304-310 (1981)); the yeast *gal4* gene sequence promoter (Johnston, S.A., *et al.*, *Proc. Natl. Acad. Sci. (USA)* 79:6971-6975 (1982); Silver, P.A., *et al.*, *Proc. Natl. Acad. Sci. (USA)* 81:5951-5955 (1984)).

As is widely known, translation of eukaryotic mRNA is initiated at the codon that encodes the first methionine. For this reason, it is preferable to ensure that the linkage between a eukaryotic promoter and a DNA sequence that encodes the *GA4H* gene (or a functional derivative thereof) does not contain any intervening codons that are capable of encoding a methionine (i.e., AUG). The presence of such codons results either in the formation of a fusion protein (if the AUG codon is in the same reading frame as the *GA4H* gene encoding DNA sequence) or a frame-shift mutation (if the AUG codon is not in the same reading frame as the *GAI* gene encoding sequence).

The *GA4H* gene encoding sequence and an operably linked promoter may be introduced into a recipient prokaryotic or eukaryotic cell either as a non-replicating DNA (or RNA) molecule, which may either be a linear molecule or, more preferably, a closed covalent circular molecule. Since such molecules are incapable of autonomous replication, the expression of the *GA4H* gene may occur through the transient expression of the introduced sequence. Alternatively, permanent expression may occur through the integration of the introduced sequence into the host chromosome.

In one embodiment, a vector is employed that is capable of integrating the desired gene sequences into the host cell chromosome. Cells that have stably integrated the introduced DNA into their chromosomes can be selected by also introducing one or more markers that allow for selection of host cells which contain the expression vector. The marker can provide for prototrophy to an auxotrophic host, biocide resistance, e.g., antibiotics, or heavy metals, such as copper, or the like. The selectable marker gene sequence can either be directly linked to the DNA gene sequences to be expressed, or introduced into the same

cell by co-transfection. Additional elements can also be needed for optimal synthesis of single chain binding protein mRNA. These elements can include splice signals, as well as transcription promoters, enhancers, and termination signals. cDNA expression vectors incorporating such elements include those described by Okayama, H., *Molec. Cell. Biol.* 3:280 (1983). In a preferred embodiment, the introduced sequence is incorporated into a plasmid or viral vector capable of autonomous replication in the recipient host. Any of a wide variety of vectors can be employed for this purpose. Factors of importance in selecting a particular plasmid or viral vector include: the ease with which recipient cells that contain the vector can be recognized and selected from those recipient cells that do not contain the vector; the number of copies of the vector that are desired in a particular host; and whether it is desirable to be able to "shuttle" the vector between host cells of different species. Preferred prokaryotic vectors include plasmids such as those capable of replication in *E. coli* (such as, for example, pBR322, ColE1, pSC101, pACYC 184,  $\pi$ VX. Such plasmids are, for example, disclosed by Maniatis, T., *et al.* (In: *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, NY (1982)). *Bacillus* plasmids include pC194, pC221, pT127, etc. Such plasmids are disclosed by Gryczan, T. (In: *The Molecular Biology of the Bacilli*, Academic Press, NY (1982), pp. 307-329). Suitable *Streptomyces* plasmids include pIJ101 (Kendall, K.J., *et al.*, *J. Bacteriol.* 169:4177-4183 (1987)), and streptomyces bacteriophages such as  $\phi$ C31 (Chater, K.F., *et al.*, In: *Sixth International Symposium on Actinomycetales Biology*, Akademiai Kiado, Budapest, Hungary (1986), pp. 45-54). *Pseudomonas* plasmids are reviewed by John, J.F., *et al.* (*Rev. Infect. Dis.* 8:693-704 (1986)), and Izaki, K. (*Jpn. J. Bacteriol.* 33:729-742 (1978)).

Preferred eukaryotic plasmids include BPV, vaccinia, SV40, 2-micron circle, etc., or their derivatives. Such plasmids are well known in the art (Botstein, D., *et al.*, *Miami Wntr. Symp.* 19:265-274 (1982); Broach, J.R., In: *The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance*, Cold

Spring Harbor Laboratory, Cold Spring Harbor, NY, p. 445-470 (1981); Broach, J.R., *Cell* 28:203-204 (1982); Bollon, D.P., *et al.*, *J. Clin. Hematol. Oncol.* 10:39-48 (1980); Maniatis, T., In: *Cell Biology: A Comprehensive Treatise*, Vol. 3, Gene sequence Expression, Academic Press, NY, pp. 563-608 (1980)).

5           Once the vector or DNA sequence containing the construct(s) has been prepared for expression, the DNA construct(s) may be introduced into an appropriate host cell by any of a variety of suitable means: transformation, transfection, conjugation, protoplast fusion, electroporation, calcium phosphate-precipitation, direct microinjection, etc. After the introduction of the vector,  
10       recipient cells are grown in a selective medium, which selects for the growth of vector-containing cells. Expression of the cloned gene sequence(s) results in the production of the *GA4H* gene, or fragments thereof. This can take place in the transformed cells as such, or following the induction of these cells to differentiate (for example, by administration of bromodeoxyuracil to neuroblastoma cells or  
15       the like).

          Following expression in an appropriate host, the GA4H protein can be readily isolated using standard techniques such as immunochromatography or HPLC to produce GA4H protein free of other *A. thaliana* proteins.

20           Genetic sequences comprising the desired gene or antisense sequence operably linked to a plant promoter may be joined to secretion signal sequences and the construct ligated into a suitable cloning vector. In general, plasmid or viral (bacteriophage) vectors containing replication and control sequences derived from species compatible with the host cell are used. The cloning vector will typically carry a replication origin, as well as specific genes that are capable of  
25       providing phenotypic selection markers in transformed host cells, typically antibiotic resistance genes.

          General methods for selecting transgenic plant cells containing a selectable marker are well known and taught, for example, by Herrera-Estrella, L. and Simpson, J. (1988) "Foreign Gene Expression in Plants" in *Plant*

*Molecular Biology, A Practical Approach*, Ed. C.H. Shaw, IRL Press, Oxford, England, pp. 131-160.

In another embodiment, the present invention relates to a transformed plant cell comprising exogenous copies of DNA (that is, copies that originated outside of the plant) encoding a *GA4* gene expressible in the plant cell wherein said plant cell is free of other foreign marker genes (preferably, other foreign selectable marker genes); a plant regenerated from the plant cell; progeny or a propagule of the plant; and seed produced by the progeny.

Plant transformation techniques are well known in the art and include direct transformation (which includes, but is not limited to: microinjection (Crossway, *Mol. Gen. Genetics* 202:179-185 (1985)), polyethylene glycol transformation (Krens *et al.*, *Nature* 296:72-74 (1982)), high velocity ballistic penetration (Klein *et al.*, *Nature* 327:70-73 (1987)), fusion of protoplasts with other entities, either minicells, cells, lysosomes, or other fusible lipid-surfaced bodies (Fraley *et al.*, *Proc. Natl. Acad. Sci. USA* 79:1859-1863 (1982)), electroporation (Fromm *et al.*, *Proc. Natl. Acad. Sci. USA* 82:5824 (1985)) and techniques set forth in U.S. Patent No. 5,231,019) and *Agrobacterium tumefaciens* mediated transformation as described herein and in (Hoekema *et al.*, *Nature* 303:179 (1983), de Framond *et al.*, *Bio/technology* 1:262 (1983), Fraley *et al.* WO84/02913, WO84/02919 and WO84/02920, Zambryski *et al.* EP 116,718, Jordan *et al.*, *Plant Cell Reports* 7:281-284 (1988), Leple *et al.* *Plant Cell Reports* 11:137-141 (1992), Stomp *et al.*, *Plant Physiol.* 92:1226-1232 (1990), and Knauf *et al.*, *Plasmid* 8:45-54 (1982), Chiang *et al.*, *Plant Cell* 7:195-201 (1995)). Another method of transformation is the leaf disc transformation technique as described by Horsch *et al.* *Science* 227:1229-1230 (1985), Bechtold *et al.*, *Acad. Sci. Paris* 316:1194-1199 (1993).

The transformation techniques can utilize DNA encoding a GA4H amino acid sequence of, including the GA4H cDNA sequence, the GA4H genomic sequence, fragments thereof or the antisense sequence, or degenerate variants of said sequences such that they are expressible in plants. Included within the scope

of a gene encoding a GA4H amino acid sequence are functional derivatives of the GA4H sequences of the invention, as well as variant, analog, species, allelic and mutational derivatives.

The preparation of functional derivatives can be achieved, for example, by site-directed mutagenesis. (Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press (1989)). Site-directed mutagenesis allows the production of a functional derivative through the use of a specific oligonucleotide that contains the desired mutated DNA sequence. One skilled in the art will recognize that the functionality of the derivative can be evaluated by routine screening assays.

As used herein, modulation of GA4H expression entails the enhancement or reduction of the naturally occurring levels of the protein. Specifically, the translation of RNA encoding GA4H can be reduced using the technique of antisense cloning.

In general, antisense cloning entails the generation of an expression module which encodes an RNA complementary (antisense) to the RNA encoding GA4H (sense). By expressing the antisense RNA in a cell which expresses the sense strand, hybridization between the two RNA species will occur resulting in the blocking of translation. Alternatively, overexpression of a GA4H protein might be accomplished by use of appropriate promoters, enhancers, and other modifications. Those of skill in the art would be aware of references describing the use of antisense genes in plants (van der Krol *et al.*, *Gene* 72:45-50 (1988); van der Krol *et al.*, *Plant Mol. Biol.* 14:467-486 (1990); Zhang *et al.*, *Plant Cell* 4:1575-1588 (1992)).

Other foreign marker genes (i.e., exogenously introduced genes) typically used include selectable markers such as a *neo* gene (Potrykus *et al.*, *Mol. Gen. Genet* 199:183-188 (1985)) which codes for kanamycin resistance; a *bar* gene which codes for bialaphos resistance; a mutant EPSP synthase gene (Hinchee *et al.*, *Bio/technology* 6:915-922 (1988)) which encodes glyphosate resistance; a nitrilase gene which confers resistance to bromoxynil (Stalker *et al.*, *J. Biol.*



*Chem.* 263:6310-6314 (1988)); a mutant acetolactate synthase gene (ALS) which confers imidazolinone or sulphonylurea resistance (EP application number 154,204); a methotrexate resistant DHFR gene (Thillet *et al.*, *J. Biol. Chem.* 263:12500-12508) and screenable markers which include  $\beta$ -glucuronidase (GUS) or an R-locus gene, alone or in combination with a C-locus gene (Ludwig *et al.*, *Proc. Natl. Acad. Sci. USA* 86:7092 (1989); Paz-Ares *et al.*, *EMBO J.* 6:3553 (1987)).

Alternatively, the genetic construct for expressing the desired protein can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. The genetic material may also be transferred into plant cells using polyethylene glycol to form a precipitation complex with the genetic material that is taken up by cells. (Paszkowski *et al.*, *EMBO J.* 3:2717-22 (1984)). The desired gene may also be introduced into plant cells by electroporation. (Fromm *et al.*, "Expression of Genes Transferred into Monocot and Dicot Plant Cells by Electroporation," *Proc. Nat'l. Acad. Sci. U.S.A.* 82:5824 (1985)). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the desired genetic construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of plasmids. Electroporated plant protoplasts reform cell walls, divide, and form plant calli. Selection of the transformed plant cells expressing the desired gene can be accomplished using phenotypic markers as described above.

Another method of introducing the desired gene into plant cells is to infect the plant cells with *Agrobacterium tumefaciens* transformed with the desired gene. Under appropriate conditions well-known in the art, transformed plant cells are grown to form shoots, roots, and develop further into plants. The desired genetic sequences can be joined to the Ti plasmid of *Agrobacterium tumefaciens*. The Ti plasmid is transmitted to plant cells on infection by *Agrobacterium tumefaciens* and is stably integrated into the plant genome. Horsch *et al.*, "Inheritance of Functional Foreign Genes in Plants," *Science* 233: 496-498 (1984); Fraley *et al.*, *Proc. Nat'l Acad. Sci. U.S.A.* 80: 4803 (1983)); Feldmann,

K.A. *et al.*, *Mol. Gen. Genet.*, 208: 1-9 (1987); Walden, R. *et al.*, *Plant J.*, 1: 281-288 (1991).

Presently there are several different ways to transform plant cells with *Agrobacterium*:

- 5 (1) co-cultivation of *Agrobacterium* with cultured, isolated protoplasts, or
- (2) transformation of cells or tissues with *Agrobacterium*.

Method (1) requires an established culture system that allows culturing protoplasts and plant regeneration from cultured protoplasts. Method (2) requires  
10 that the plant cells or tissues can be transformed by *Agrobacterium* and that the transformed cells or tissues can be induced to regenerate into whole plants. In the binary system, to have infection, two plasmids are needed: a T-DNA containing plasmid and a *vir* plasmid.

Routinely, however, one of the simplest methods of plant transformation  
15 is explant inoculation, which involves incubation of sectioned tissue with *Agrobacterium* containing the appropriate transformation vector (Plant Genetic Transformation and Gene Expression, A Laboratory Manual, Oxford: Blackwell Scientific Publications (1988); Walden, Genetic Transformation in Plants, Milton Koynes: Open University Press (1988)).

20 All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be used for the expression of the desired gene. Suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manicot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*,  
25 *Datura*, *Hyoscyamus*, *Lycopersion*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hemerocallis*, *Nemesia*, *Pelargonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browallia*, *Glycine*, *Lolium*, *Zea*, *Triticum*, *Sorghum*, and *Datura*. Additional plant genera that may be transformed by  
30 *Agrobacterium* include *Ipomoea*, *Passiflora*, *Cyclamen*, *Malus*, *Prunus*, *Rosa*,

*Rubus*, *Populus*, *Santalum*, *Allium*, *Lilium*, *Narcissus*, *Ananas*, *Arachis*, *Phaseolus*, and *Pisum*.

Plant regeneration techniques are well known in the art and include those set forth in the *Handbook of Plant Cell Culture*, Volumes 1-3, Eds. Evans *et al.* Macmillan Publishing Co., New York, NY (1983, 1984, 1984, respectively);  
 5 Predieri and Malavasi, *Plant Cell, Tissue, and Organ Culture* 17:133-142 (1989); James, D.J., *et al.*, *J. Plant Physiol.* 132:148-154 (1988); Fasolo, F., *et al.*, *Plant Cell, Tissue, and Organ Culture* 16:75-87 (1989); Valobra and James, *Plant Cell, Tissue, and Organ Culture* 21:51-54 (1990); Srivastava, P.S., *et al.*, *Plant Science*  
 10 42:209-214 (1985); Rowland and Ogden, *Hort. Science* 27:1127-1129 (1992); Park and Son, *Plant Cell, Tissue, and Organ Culture* 15:95-105 (1988); Noh and Minocha, *Plant Cell Reports* 5:464-467 (1986); Brand and Lineberger, *Plant Science* 57:173-179 (1988); Bozhkov, P.V., *et al.*, *Plant Cell Reports* 11:386-389 (1992); Kvaalen and von Arnold, *Plant Cell, Tissue, and Organ Culture* 27:49-57  
 15 (1991); Tremblay and Tremblay, *Plant Cell, Tissue, and Organ Culture* 27:95-103 (1991); Gupta and Pullman, U.S. Patent No. 5,036,007; Michler and Bauer, *Plant Science* 77:111-118 (1991); Wetzstein, H.Y., *et al.*, *Plant Science* 64:193-201 (1989); McGranahan, G.H., *et al.*, *Bio/Technology* 6:800-804 (1988); Gingas, V.M., *Hort. Science* 26:1217-1218 (1991); Chalupa, V., *Plant Cell Reports*  
 20 9:398-401 (1990); Gingas and Lineberger, *Plant Cell, Tissue, and Organ Culture* 17:191-203 (1989); Bureno, M.A., *et al.*, *Phys. Plant.* 85:30-34 (1992); and Roberts, D.R., *et al.*, *Can. J. Bot.* 68:1086-1090 (1990).

Plant regeneration from cultured protoplasts is described in Evans *et al.*, "Protoplast Isolation and Culture," in *Handbook of Plant Cell Culture* 1:124-176  
 25 (MacMillan Publishing Co., New York, 1983); M.R. Davey, "Recent Developments in the Culture and Regeneration of Plant Protoplasts," *Protoplasts, 1983 - Lecture Proceedings*, pp. 19-29 (Birkhauser, Basel, 1983); P.J. Dale, "Protoplast Culture and Plant Regeneration of Cereals and Other Recalcitrant Crops," in *Protoplasts 1983 - Lecture Proceedings*, pp. 31-41 (Birkhauser, Basel,

1983); and H. Binding, "Regeneration of Plants," in *Plant Protoplasts*, pp. 21-37 (CRC Press, Boca Raton, 1985).

Techniques for the regeneration of plants varies from species to species but generally, a suspension of transformed protoplasts containing multiple copies of the desired gene is first provided. Embryo formation can then be induced from the protoplast suspensions, to the stage of ripening and germination as natural embryos. The culture media will generally contain various amino acids and hormones, such as auxins and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa.

Mature plants, grown from transformed plant cells, are selfed to produce an inbred plant. The inbred plant produces seed containing the recombinant DNA sequences promoting increased expression of GA4H.

Parts obtained from regenerated plants, such as flowers, seeds, leaves, branches, fruit, and the like are covered by the invention provided that these parts comprise the herbicidal tolerant cells. Progeny and variants, and mutants of the regenerated plants are also included within the scope of this invention. As used herein, variant describes phenotypic changes that are stable and heritable, including heritable variation that is sexually transmitted to progeny of plants, provided that the variant still comprises a herbicidal tolerant plant through enhanced rate of acetylation. Also, as used herein, mutant describes variation as a result of environmental conditions, such as radiation, or as a result of genetic variation in which a trait is transmitted meiotically according to well-established laws of inheritance.

Plants which contain the GA4H encoding DNA of the invention and no other foreign marker gene are advantageous in that removal of the foreign marker gene, once inserted into the plant, may be impossible without also removing the GA4H gene. Absence of the foreign marker gene is sometimes desired so as to minimize the number of foreign genes expressed. This can be achieved by providing the GA4H-encoding DNA between Ti-plasmid borders.

The *GA4H* gene product may have similar function(s) to 3- $\beta$ -hydroxylase. 3- $\beta$ -hydroxylase is critical for controlling stem growth (Ingram *et al.*, *Plant* 160: 455-463 (1984). Accordingly, the GA4H of the invention may be applied to crops to enhance and facilitate such stem elongation, flowering and fruiting. Alternatively, the DNA encoding GA4H may be genetically inserted into the plant host to produce a similar effect.

All plants which can be transformed are intended to be hosts included within the scope of the invention (preferably, dicotyledonous plants). Such plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersicon*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hererocallis*, *Nemesia*, *Pelargonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Sencia*, *Salpiglossis*, *Cucumis*, *Browalia*, *Glycine*, *Lolium*, *Zea*, *Triticum*, *Sorghum*, *Malus*, *Apium*, *Datura*, the *le* mutant in peas, the *ga4* mutant in *Arabidopsis*, and the *dwarf-1* mutant in *Monocotyledonous* plants such as corn.

Examples of commercially useful agricultural plants useful in the methods of the invention as transgenic hosts containing the GA4 DNA or antisense sequence of the invention include grains, legumes, vegetables and fruits, including but not limited to soybean, wheat, corn, barley, alfalfa, cotton, rapeseed, rice, tobacco, rye, tomatoes, beans, peas, celery, grapes, cabbage, oilseed, apples, strawberries, mulberries, potatoes, cranberries and lettuce.

Having now generally described the invention, the same will be more readily understood through reference to the following examples which are provided by way of illustration, and are not intended to be limiting of the present invention, unless specified.

## *Examples*

### *Example 1*

#### *Isolation of The GA4 Homologue Genes*

5 The presence of a *GA4*-homologue gene (*GA4H*) was first determined by low stringency hybridization using a probe made from the *GA4* sequence. The probe was designed based on the DNA sequence of a conserved amino acid region between *GA4* and similar proteins (i.e.  $\beta$ -hydroxylases).

#### *Methods*

##### *"Plant and Nucleic Acid Sources and Preparation"*

10 A *ga4-1* (an ethyl methanesulfonate, EMS, induced mutant) mutant was obtained from M Koornneef (Agricultural University, Wageningen, The Netherlands). Plants were grown under a 16-hr light/ 8-hr dark cycle. For genomic DNA isolation, rosette leaves of 3-4 week old plants were harvested and frozen in liquid nitrogen. For RNA isolation, tissues from matured flowering plants of  
15 either *ga4-1* or *Lansberg erecta* were collected and immediately frozen in liquid nitrogen.

pCD7 DNA containing the *GA4* cDNA has been described previously (Chiang, H.H., *et al.*, *Plant Cell* 7:195-201 (1995)). The cloning vectors were either pBSKS(-) or pBSKS(+) of Stratagene (La Jolla, CA, U.S.A.). DNA  
20 markers, 1 Kb and 123 bp, are from Gibco BRL (Gaithersburg, MD, U.S.A.). Restriction and modifying enzymes were from New England Biolab (Cambridge, MA, U.S.A.).

Genomic DNA of yeast strains carrying YAC DNA was isolated according to Ausubel, F.M., *et al.*, *Current Protocols in Molecular Biology*, New  
25 York: Greene Publishing Association and Wiley-Interscience (1987). Plant genomic DNA was isolated by the method of Watson, J.C., and Thompson, W.F., *Methods in Enzymology* 118:57-75 (1986). RNA was isolated using the Tri-Reagent (Molecular Research Center, Cincinnati, OH, U.S.A.).

"Oligonucleotides and Sequence Analysis"

Oligonucleotides were synthesized by the DNA Synthesis Core Facility of the Molecular Biology/ Endocrine Departments of Massachusetts General Hospital (MGH) (Boston, MA, U.S.A.). In the following oligonucleotides the underlined nucleotides indicate the restriction recognition site shown in parenthesis. The name and sequence of the oligonucleotides are as follows:

Homo1: 5'-GTGGTTAGCACTAAATTCAC-3' (SEQ ID No. 11)

Homo2: 5'-GACCCATGGCTCGGTCCGGT-3' (SEQ ID No. 12)

GA-P1X: 5'-GCTCTAGAGAGTATTTGAGAAGG-3' (SEQ ID No. 13)  
(*Xba*I)

GA-P2: 5'-GTTTACTATTGCCGATGACT-3' (SEQ ID No. 14)

GA-P6: 5'-CAATACCAAAAATGAAAAGC-3' (SEQ ID No. 15)

GA-P13: 5'-CTCCTACCGCAACCATTTC-3' (SEQ ID No. 16)

GA-P14S: 5'-TCCCCCGGGTTTATGTGATGAGCATCCC-3' (SEQ ID No. 17)

(*Sma*I)

GA-P15: 5'-CCAAAGTAATTGTTTATGTG-3' (SEQ ID No. 18)

GA-P16: 5'-AATTTAGGTTTTTCATTAAG-3' (SEQ ID No. 19)

GA-P17: 5'-GTAGTGGTTTAGTCGTATGG-3' (SEQ ID No. 20)

GA-P18: 5'-AAAACCTGGAGACCGGCGG-3' (SEQ ID No. 21)

GA-P19: 5'-TATCATGTAATCTTTTTGG-3' (SEQ ID No. 22)

GA-P20: 5'-CCGGCTTCCCGTACAGCGG-3' (SEQ ID No. 23)

GA-P21: 5'-AATCAAGAAATTCAGTCGG-3' (SEQ ID No. 24)

GA-P27E: 5'-GGAATTCATACCAAAAACATAAAGCC-3' (SEQ ID No. 25)  
(*Eco*RI)

Tua4F: 5'-CTAGTTTCTTTCTTCCACG-3' (SEQ ID No. 26)

Tua4R: 5'-TAGCTGCATCTTCTTTACC-3' (SEQ ID No. 27)

DNA sequences were determined by the DNA Sequencing Core Facility of the Department of Molecular Biology at Massachusetts General Hospital.

Sequence analyses were performed using the software package of the Genetics Computer Group (GCG; Madison, WI, U.S.A.). Blast searches were conducted through the National Center for Biotechnology Information (NCBI), (Bethesda, MD, U.S.A.) using the algorithm of Altschul, S.F., *et al.*, *J. Mol. Biol.* 215:403-10 (1990).

#### "Polymerase Chain Reaction"

PCR was performed using the Peltier Thermal Cycler (PTC-200) of MJ Research (Watertown, MA, U.S.A.). A DNA fragment containing a conserved region on the second exon of the GA4 gene (Chiang, H.H., *et al.*, *Plant Cell* 7:195-201 (1995)) was generated by PCR using Homo1 and Homo2 primers. Probes prepared from this fragment (Homologous probes) were used for the genomic DNA gel blot and for screening the genomic library. The PCR reaction was carried out in 100 µl total volume and contained 0.4 ng of pCD7 DNA, 200 µM of dNTP, 15 µM of each primer, and 2.5 units of *Taq* DNA polymerase (Boehringer Mannheim, Indianapolis, IN, U.S.A.). The PCR temperature profile was 35 cycles of 1 minute at 94°C, 1 minute at 50°C, and 3 minutes at 72°C. Preparation of the Unique probes were described earlier (Chiang, H.H., *et al.*, *Plant Cell* 7:195-201 (1995)).

In the mapping study, 1 µg of each YAC DNA was used as templates for PCR amplification of the two homologous genes. The GA4H1 gene was amplified using GA-P2 and GA-P6 primers. The GA4H2 gene was amplified using GA-P19 and GA-P20 primers. Each PCR reaction was carried out in 25 µl total volume and contained 80 µM of dNTPs, 10 µM of each primer, and 2 units of *Taq* DNA polymerase (Boehringer Mannheim). The PCR was performed using 35 cycles of 40 seconds at 92°C, 40 seconds at 55°C, and 40 seconds at 72°C. One fifth of the PCR product was separated on 0.8% agarose gel.

#### "RT-PCR Conditions"

First strand cDNA synthesis was performed according to Sambrook, J.,



*et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor: Cold Spring Harbor Laboratory (1989). The reaction volume was 25  $\mu$ l and it contained 1  $\mu$ g of total RNA, 9  $\mu$ M of (dT)<sub>20</sub>, 1.2 mM dNTP, 136 units of RNASE inhibitor (Amersham, Arlington Heights, IL, U.S.A.), and 9.5 units of avian myeloblastosis virus (AMV) reverse transcriptase (Promega, Madison, WI, U.S.A.). The reaction was incubated at 42°C for one hour and then at 72°C for 15 minutes. Eight microliters of the first strand cDNA was used as templates in the PCR amplification. The reaction was in 50  $\mu$ l and used 63  $\mu$ M of dNTP, 0.6  $\mu$ M of homologous gene specific primer, 0.4  $\mu$ M of tubulin primer, and 2.5 units of *Taq* DNA polymerase (Boehringer Mannheim). The thermal profile was 40 cycles of 45 seconds at 94°C, 45 seconds at 55°C, and 45 seconds at 72°C. When amplifying the full length cDNA, tubulin primers were not included and the extension time of 45 seconds at 72°C was increased to 1.5 minutes. One-tenth of the PCR product was analyzed on an agarose gel.

#### "Genomic Library Screening"

An *Arabidopsis* genomic library made from ecotype C24 was kindly provided by Dr. Lin Sun (Nemapharm, Cambridge, MA, U.S.A.). This library was constructed using the *Sau*3A partial digested genomic DNA and subsequently cloned into the *Xho*I site the  $\lambda$ FIX-II vector (Stratagene). Screening of the library was performed according to the manufacturer's protocol (Stratagene). Plaques were transferred and crosslinked to Biotrans nylon membrane by autoclaving for 2 minutes. Homologous probes was prepared and the hybridization conditions were as described in Chiang, H.H., *et al.*, *Plant Cell* 7:195-201 (1995), except that Homo1 and Homo2 primers were used and filters were hybridized at 42°C (low stringency). Filters were washed once in 2X SSC (1X= 0.15 M NaCl, 0.015 M sodium citrate) for 15 minutes at room temperature and twice in 0.1X SSC, 0.1% SDS for 30 minutes at 42°C (low stringency).

### "DNA Gel Blot Analysis"

In the genomic Southern, *Arabidopsis* (ecotype Landsberg *erecta*) genomic DNA was digested with appropriate restriction enzymes, separated by agarose gel electrophoresis, and transferred to Biotrans membrane (ICN Biomedical Inc., Aurora, OH USA) as described in Chiang, H.H., *et al.*, *Plant Cell* 7:195-201 (1995). For the homologous and unique GA4 gene probes, the hybridization and washing conditions were the same as the library screening above (low stringency). The DNA gel blot analysis using the GA4H1 gene, p3-1, probes was performed as described in Chiang, H.H., *et al.*, *Plant Cell* 7:195-201 (1995). The hybridization and washing conditions were performed at 65°C (high stringency).

DNA blot analyses for the mapping and RT-PCR products were performed as described (Cheng, C.L., *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89:1861-4 (1992)). In the mapping of homologous genes by PCR, probes specific to these genes were generated by PCR. Probes were prepared using a 4.4 kbp *Bgl*II/ *Xho*I genomic DNA fragment, containing these two genes, as templates with four primers (GA-P2, GA-P6, GA-P19, and GA-P20). The reaction was in 50 µl, and it contained 5 ng of DNA template, 100 µM each of dCTP, dGTP, and dTTP, 5 µM dATP, 50 µCuries of  $\alpha$ -<sup>32</sup>P dATP (Dupont NEN, Wilmington, DE, U.S.A.) 0.4 µM each primer, and 2.5 units of *Taq* DNA polymerase (Boehringer Mannheim). The thermal profile was 30 cycles of 40 seconds at 94°C, 30 seconds at 55°C, and 30 seconds at 72°C.

In the RT-PCR DNA gel blot, the same PCR method as above was used to prepare the GA4H1 and GA4H2 specific probes, except that different primers were employed. Primers pairs of GA-P13/ GA-P17 and GA-P18/ GA-P20 were used to prepare GA4H1 and GA4H2 gene probes, respectively.

### Results

To isolate the DNA sequences with similar sequence to the GA4 gene (ATCC accession nos. 98393 and 98394), low stringency hybridization (see Materials and Methods) to *Arabidopsis* genomic DNA was performed with

Homologous probes (SEQ ID No. 2) prepared from a conserved region the *GA4* gene (Figure 1), compared to *GA5* and other  $\beta$ -hydroxylases. Results from the blot of this genomic DNA, isolated from ecotype *Lansberg erecta*, are shown in Figure 2A.

5           Beside a strong 3.2 kbp size band in the *Hind*III digested DNA, a less intense 2.1 kbp band is visible and assumed to contain DNA similar to the *GA4* gene (Figure 2A, lane 1). Similarly, there is a light 2.8 kbp band in the *Bam*HI digested DNA (Figure 2A, lane 2).

10           To identify the *GA4* gene, a similar blot was hybridized at low stringency to a Unique probe (Figure 1 - SEQ ID No. 3) derived from a less conserved region of the *GA4* gene. This probe would hybridize specifically to the *GA4* gene, and results are shown in Figure 2B.

15           In the *Hind*III digestion, the *GA4* specific probes hybridized strongly to the 3.2 kbp size band, and no detectable signal was found at the 2.1 kbp size (Figure 2B, lane 1). Similarly, the 2.8 kbp band in the *Bam*HI digested DNA was not visible, indicating that the 2.1 kbp *Hind*III and the 2.8 kbp *Bam*HI fragments contain a homologous sequence to the *GA4* DNA (Figure 2B, lane 2). DNA digested with the *Eco*RI enzyme resulted in only high molecular weight bands being visible when either Homologous or Unique probes were used (Figure 2A and 2B, lane 3).

20           The homologous probes were also used to screen a genomic library (ecotype C24) at low stringency conditions as described above. In addition to the *GA4* genomic clones, one other genomic clone ( $\lambda$ 3) that contained the 2.1 kbp *Hind*III fragment was isolated. This 2.1 kbp fragment of  $\lambda$ 3 was subcloned into pBSKS(-) to produce p3-1 (Figure 3). The whole genomic insert in  $\lambda$ 3 was also  
25           cloned into pBSKS(+) using the *Not*I sites flanking the insert to generate pLVN103. To confirm this genomic clone, the p3-1 DNA was used as a probe and hybridized at high stringency to the same genomic blot above. As shown in Figure 2C, both the 2.1 kbp *Hind*III and 2.8 kbp *Bam*HI fragments are present  
30           (lane 1 and 2). The predicted high molecular weight fragment in *Eco*RI digested

-38-

DNA is also present (lane 3). These results indicated that the predicted homologue of the *GA4* gene had been isolated.

Clones p3-1 and part of pLVN103 DNAs were sequenced, and the homologue gene was named GA4H1. Further sequencing in the 5' flanking of the GA4H1 gene revealed a second gene, named GA4H2, that also has sequence similarity to the GA4 as well as to the GA4H1 genes. The genome organization of these two linked genes is represented in Figure 3. When compared to the GA4 gene, both the GA4H1 and GA4H2 genes also possess a single intron that is located at a similar position in the gene. Transcription of both genes is in the same direction, and they are separated by a 1 kbp spacer region (Figure 3).

The plasmid designated pLVN103 comprising the genomic sequence of both the GA4H1 and GA4H2 genes was deposited at the ATCC (Rockville, MD.) under the terms of the Budapest Treaty and has been granted accession number 98436.

### ***Example 2*** ***Chromosomal Location of the Homologue Genes***

It was determined that both homologue genes are located on chromosome 1. Since many continuous overlapping DNA clones of Yeast Artificial Chromosomes (YAC) containing *Arabidopsis* genomic DNA had been placed on the five linkage groups, the GA4H1 and GA4H2 genes can be mapped by anchoring them to YACs of known position.

Probes derived from the genomic clone p3-1 were hybridized to the CIC YAC library (Creusot, F., *et al.*, *Plant Journal* 8:763-70 (1995)), and three YAC clones (CIC1E4, CIC6C10 and CIC10A11) were isolated (data not shown). The intensity of the hybridization was higher in CIC1E4 and CIC6C10 than in CIC10A11 (data not shown). These putative YACs were subsequently confirmed by PCR amplification using primers specific to these two genes.

Two specific primer sets (GA-P2/GA-P6 and GA-P19/GA-P20 for GA4H1 and GA4H2 genes, respectively) were used to amplify a short region in

these genes. The predicted amplified products for *GA4H1* and *GA4H2* genes are 480 bp and 410 bp, respectively. The analysis of PCR products is shown in Figure 4A. For the *GA4H2* gene, the predicted PCR product of 410 bp was present in both the control pLVN103 DNA (lane 2) and in two of the three putative YACs, CIC1E4 (lane 4) and CIC6C10 (lane 5). However, CIC10A11 YAC did not appear to carry the *GA4H2* gene, since the 410 bp size band was not present (lane 6).

The CIC6C3 YAC, located on the bottom of chromosome 2, was used as a negative control. As expected, no PCR product was present in CIC6C3, indicating the specificity of these primers (lane 1). Similar results were also obtained for the *GA4H1* gene where the predicted PCR product is 480 bp in size. The 480 bp size band was present in the pLVN103 control (lane 8) as well as in CIC1E4 and 6C10 (lanes 10 and 11). Again, the 480 bp size band was absent in CIC10A11. These results were further confirmed by the DNA gel blot.

Probes, generated using the same 4 primers with the genomic clone (pLVN103), were hybridized to the DNA blot, and the results are shown in Figure 4B. All predicted PCR products of 410 bp and 480 bp in size (for *GA4H2* and *GA4H1* genes, respectively) were hybridized to the probes. Since both CIC1E4 and CIC6C10 were previously anchored to the bottom of chromosome 1, it was concluded that *GA4H1* and *GA4H2* genes are located at about 159-cM (on the physical map) of chromosome 1 (<http://cbil.humgen.upenn.edu/~atgc/ATGCUP.html>; [http://cbil.humgen.upenn.edu/~atgc/physical-mapping/xlchl\\_pt4.html](http://cbil.humgen.upenn.edu/~atgc/physical-mapping/xlchl_pt4.html)). CIC10A11 has overlapping regions to those two YACs above, and it hybridized weakly to probes prepared from p3-1. However, no PCR product was amplified when CIC10A11 was used as a template DNA. These results suggest that the edge of CIC10A11 DNA may end shortly after the *HindIII* site, located in the 3' flanking of the *GA4H1* gene (see Figure 3).

**Example 3**  
**Cloning of GA4H1 and GA4H2 cDNAs By RT-PCR**

To determine whether the GA4H1 gene is expressed, probes derived from the clone p3-1 containing most of the GA4H1 coding region were used to hybridize to RNA isolated from flowers, shoot meristems, leaves, roots and siliques. However, no visible signal was present in the RNA blot (data not shown). Another attempt to isolate the cDNA by screening a yeast expression library (Minet, M., *et al.*, *Plant Journal* 2:417-422 (1992)) using probes derived from p3-1 also failed. Furthermore, searching the *Arabidopsis* EST database using the GA4H1 sequence no match was found to any known EST, indicating that the *GA4H1* gene may be expressed at very low levels or only in a specific developmental stage of the plant. Therefore, isolation of the GA4H1 cDNA by reverse-transcriptase PCR (RT-PCR) was undertaken.

The *ga4* mutant was used as a source of RNA since the expression of the *GA4* gene is under feedback regulation resulting in the induction of its mRNA (Chiang, H.H., *et al.*, *Plant Cell* 7:195-201 (1995)). If the expression of the *GA4H1* gene is regulated by the same or a similar mechanism, i.e. a higher level of GA4H1 mRNA in the *ga4* mutant than wild type, then one has a better chance of obtaining the cDNA in the *ga4* mutant background.

RT-PCR was performed using RNAs isolated from whole seedlings of *ga4-1* (EMS) and *ga4-2* (T-DNA) mutants grown in liquid and from leaves and inflorescences of soil grown *ga4-1* plants. Inflorescences contain the shoot meristems, flowers and siliques. A predicted PCR product was observed only in RNA isolated from inflorescence tissues (data not shown). Therefore, inflorescences were used as a source of RNA for cloning the GA4H1 cDNA. Primers GA-P15 and GA-P16 were used in PCR following the reverse transcription. A nested PCR using GA-P1X and GA-P14S primers was performed, and the product was subsequently cloned into pBSKS(+) at the *Sma*I and *Xba*I sites.

Since *Taq* DNA polymerase, a low fidelity enzyme, was used in the PCR amplification, three independent RT-PCR clones (pLVN107a, b, c) were sequenced. The consensus sequence of this cDNA clone, labeled as pLVN107, is shown in Figure 5 (SEQ ID No. 5).

5           The cDNA contains 43 and 22 nucleotides in the 5' and 3' untranslated regions of the gene, respectively. Four of the nine nucleotides in the sequence surrounding the predicted start codon (ATG) are identical to the consensus sequence (Joshi, C. P., *Nucleic Acids Res.* 15:6643-53 (1987)). The intron occurs at a similar position relative to the GA4 gene. The GA4H1 genomic DNA  
10           sequence (SEQ ID No. 6), along with its deduced amino acid sequence (SEQ ID No. 7), are shown in Figure 6. The gene possesses a single 409 bp intron, and it follows the intron's GT/AG consensus rule. This gene encodes a protein of 355 amino acids long.

          Comparison between the RT-PCR sequence (pLVN107) and the genomic  
15           sequence (pLVN103) revealed one nucleotide mismatch at the position no. 1059 of the cDNA sequence (Figure 5) (SEQ ID No. 5). The cDNA has a "G" at this position while the genomic DNA has an "A". This mismatch may arise from differences in the *Lansberg erecta* (L. *er.*) and C24 ecotypes from which cDNA and genomic sequences were derived, respectively.

20           To resolve this, the genomic DNA of the L. *er.* ecotype was cloned by PCR amplification with a high fidelity enzyme, *Pfu* (Stratagene), using GA-P1X and GA-P14S primers. The sequence of this clone, pLVN110, is identical to the genomic clone in C24 ecotype, pLVN103 (data not shown). Therefore, the mismatch at this position could not be resolved by current data.

25           Similar RT-PCR conditions were used to isolate the GA4H2 cDNA except that GA-P27E and GA-P21 primers (SEQ ID Nos. 11 and 14 respectively) were used, and the RNA source was of *Lansberg erecta*. One cDNA was cloned, pLVN115, and its sequence (SEQ ID No. 8) is shown in Figure 7.

30           Similar to GA4 and GA4H1 gene, there is a single intron present at a conserved position in the gene. The sequence surrounding the predicted ATG

show 3/9 matches against the consensus sequence. The genomic sequence of the gene is shown in Figure 8 (SEQ ID No. 9). Sequence comparison between this cDNA and its genomic DNA shows a perfect match. The GA4H2 gene encodes a protein of 347 amino acids long (SEQ ID No. 10).

5

**Example 4**  
***Sequence Analysis of the GA4H1 and GA4H2 Proteins***

Each of the protein sequences (SEQ ID Nos. 7 and 10) of the GA4H1 and GA4H2 genes was searched against the protein Genbank database, and each time the GA4 protein was found to be the best match. This is not surprising, since these genes were isolated by hybridization to probes prepared from the GA4 DNA sequence.

The predicted proteins encoded by the GA4, GA4H1, and GA4H2 genes were compared and the results of this comparison are shown in Figure 9. As expected, many conserved regions are present throughout these proteins. However, GA4H2 protein has higher homology to GA4 than does GA4H1. Amino acid sequence identity was calculated among these proteins using GAP software of the GCG package and the results are shown in Figure 10. GA4H2 and GA4 share 76% and 85% amino acid identity and similarity, respectively. Compared to this, GA4H1 and GA4 only share 57% and 73% amino acid identity and similarity, respectively. Results of comparison between GA4H1 and GA4H2 are similar to those between GA4H1 and GA4.

Several enzymes, including GA4 ( $\beta$ -hydroxylase), in the gibberellin biosynthesis pathway belong to a group of non-heme iron-containing enzymes called 2-oxoacid-dependent dioxygenases (2-ODD). A binary comparison between these proteins and the three proteins described above is shown in Figure 10. Proteins of different functions often share around 30% amino acid identity, while those from a multigene family in the same species show greater than 50% amino acid identity (Prescott, A.G., and John, P., *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47:245-271 (1996)). Results in Figure 10 appear to support this



observation with the exception of the GA4, GA4H1 and GA4H2. These three proteins share greater than 50% amino acid identity, which indicates that they belong to the same family and/ or may have similar enzyme activities.

**Example 5**  
**Differential Expression of GA4H1 and GA4H2 Genes**

Since the expression of the *GA4* gene is primarily in the silique, the expression levels of *GA4H1* and *GA4H2* genes in various organs was investigated to determine whether a similar expression pattern occurred. RT-PCR using *Arabidopsis* (Lansberg *et al.*) RNAs isolated from liquid grown roots, soil grown rosette leaves, floral shoots (including flowers), and siliques was performed. GA-P13/ GA-P17 and GA-P18/ GA-P20 primer pairs were used to amplify the GA4H1 and GA4H2 genes, respectively. Primers in each pair, located on separated exons were used to differentiate between cDNA and genomic DNA. The predicted RT-PCR products of GA4H1 and GA4H2 genes are 220 bp and 440 bp, respectively. The predicted PCR products of GA4H1 and GA4H2 genomic DNAs (containing the intron sequence) are 630 bp and 860 bp, respectively.

Primers from the  $\alpha$ -tubulin 4 gene, *TUA4* (Kopczak, S.D., *et al.*, *Plant Cell* 4:539-47 (1992)), were used as an internal control along with GA4 homologue gene specific primers. The  $\alpha$ -tubulin primers generates a 320 bp RT-PCR product. Results of RT-PCR analysis are shown in Figure 11A.

To confirm the PCR products, a DNA gel blot analysis was performed using probes derived from the *GA4H1* gene (Figure 11B). The *GA4H1* gene was mainly expressed in the flowers and shoot meristems, with smaller amounts in the siliques (Figures 11 A and 11B, lanes 5 and 8). In addition, *GA4H1* gene was barely detected in the root tissues (Figures 11A and 11B, lane 7). However, there was no detectable level of GA4H1 gene in the rosette leaves (Figures 11A and 11B, lane 6).

Similar to the polymerase chain reaction control, the 630 bp product was present in pLVN103 containing the genomic clone (lane 3). There was a small amount of genomic DNA present in the RNA preparation, as indicated by the presence of the 630 bp size band in all tissue types. pCD7 (GA4 cDNA clone) and pLVN115 (GA4H2 cDNA clone) were also used as templates to demonstrate the specificity of GA-P13 and GA-P17 primer pair.

Although some unspecific PCR products (Figure 11A, lanes 1 and 2) were present, these primers amplified neither the *GA4* nor the *GA4H2* gene (Figure 11B, lanes 1 and 2). The internal RT-PCR control ( $\alpha$ -tubulin 4 gene) was present evenly in different tissue types with the exception of siliques (Figure 11A, lane 8). This may indicate that less silique RNA was used in this experiment, suggesting that the expression level of these genes in siliques was underestimated.

A similar experiment was performed on the *GA4H2* gene where GA-P18 and GA-P20 primers were used in the amplification. Again, there was less silique RNA used, as indicated in Figure 12A (lanes 1-4). Unlike the *GA4H1* gene, *GA4H2* transcripts were more abundant in the root tissues, while lower levels were present in the flowers and shoot meristems (Figure 11A and 11B, lanes 2 and 4). In addition, *GA4H2* expression is barely detected in siliques but not in leaves (Figure 11A and 11B, lane 1 and 3). Again, the expression level of *GA4H2* gene in siliques was underestimated when compared to other tissues. A genomic DNA clone (pLVN103) was used as the control, and it possess the predicted 860 bp size band (Figure 11A and 11B, lane 6). Similar to the *GA4H1* RT-PCR result, primers used in this experiment were specific to the *GA4H2* gene (Figure 11A and 11B, lanes 7 and 8).

### Example 6

#### *Expression of Antisense GA4H RNA*

An expression vector is constructed using methods well known in the art, such that it expresses an RNA complementary to the sense strand GA4H RNA. The antisense GA4H RNA is expressed in a constitutive fashion using promoters

that are constitutively expressed in a given host plant, for example, the cauliflower mosaic virus 35S promoter. Alternatively, the antisense RNA is expressed in a tissue specific fashion using tissue specific promoters. As described earlier, such promoters are well known in the art.

5 In one example, the antisense construct pPO35 (Oeller *et al.*, *Science* 254:437-439 (1991)) is cut with BamH1 and Sac1 to remove the tACC2 cDNA sequence. After removing the tACC2 cDNA, the vector is treated with the Klenow fragment of *E. coli* DNA polymerase I to fill in the ends, and the sequence described in Figure 6 or 8 is blunt end ligated into the vector such that  
10 the strand operably linked to the promoter is that which transcribes the GA1 antisense RNA sequence. The ligated vector is used to transform an appropriate *E. coli* strain.

Colonies containing the ligated vector are screened using colony hybridization or Southern blotting to obtain vectors which contain the *GA4H*  
15 cDNA in the orientation which will produce antisense RNA when transcribed from the 35S promoter contained in the vector.

The antisense GA4H vector is isolated from a colony identified as having the proper orientation and the DNA is introduced into plant cells by one of the techniques described earlier, for example, electroporation or *Agrobacterium*/Ti  
20 plasmid mediated transformation.

Plants regenerated from the transformed cells express antisense GA4H RNA. The expressed antisense GA4H RNA binds to sense strand GA4H RNA and thus prevents translation.

In an initial experiment the phenotypes of transgenic plants expressing the  
25 antisense of the GA4H1 gene were examined. Constructs carrying the sense and antisense of the GA4H1 cDNA, under transcriptional control of the cauliflower mosaic virus 35S promoter, were transferred into *Arabidopsis thaliana* ecotype Landsberg *erecta* via *Agrobacterium* mediated transformation (Bechtold *et al.*, *Acad. Sci. Paris* 316:1194-1199 (1993)). These constructs contained a neomycin phosphotransferase (NPT-II) gene whose product confers resistance to kanamycin.  
30

Transgenic seed were harvested and subsequently germinated on MS medium supplemented with 50 mg/L kanamycin. Resistant seedlings (T1 generation) were transplanted to soil and the height was measured on mature plants. Untransformed plants, *Lansberg erecta* ecotype, were grown similarly but in the absence of kanamycin.

Results of transgenic plants carrying the sense or antisense cDNA of the *GA4H1* gene are shown in Figure 13. Overexpression of the *GA4H1* cDNA in the sense orientation does not seem to alter the plant's height. However, several plants carrying the antisense of the *GA4H1* cDNA exhibit dwarf phenotype. These preliminary results require further validation, especially in the subsequent generation. These results suggest that one can use the *GA4H1* gene in the antisense orientation to generate dwarf plants.

#### **EXAMPLE 7** ***GA4H Protein Level in Wild-Type and Transgenic Lines***

##### ***Agrobacterium tumefaciens-mediated transformation of Arabidopsis root explants.***

The transformation procedure is described previously (Valvekens *et al.*, 1988) with slight modifications (Sun *et al.*, *Plant Cell* 4:119-128 (1992)). Sense or anti-sense DNA is introduced into *Agrobacterium* LBA4404 by electroporation (Ausubel *et al.*, *Current Protocols in Molecular Biology* (New York: Green Publishing Associates/Wiley-Interscience) (1990). Stability of the insert of the plasmid in LBA4404 is tested by restriction digestion and gel electrophoresis of plasmid DNA purified by NaOH/SDS miniprep procedure (Ausubel *et al.*, *Current Protocols in Molecular Biology* (New York: Green Publishing Associates/Wiley-Interscience) (1990).

A fresh overnight culture of LBA4404 carrying individual plasmids is used to infect root explants of four-week-old wild-type plants. Km<sup>r</sup> transgenic plants are regenerated as described (Valvekens *et al.*, *Proc. Natl. Acad. Sci. USA* 85:5536-5540 (1988)). Seeds of transgenic plants are germinated on MS agar

plates containing kanamycin (50 µg/ml). Non-germinating seeds after 8 days were transferred onto MS plates containing 100 µM GA<sub>3</sub> and 50 µg/ml kanamycin to score for GA<sup>+</sup>/Km<sup>r</sup> and GA<sup>-</sup>/Km<sup>s</sup> segregation.

5 The levels of GA4H proteins in both sense and antisense transgenic *Arabidopsis* plants are compared to the level in wild-type plants (ecotype Landsberg *erecta*) by immunoblot analysis. Supernatant fractions, are obtained by tissue extraction and centrifugation (Bensen and Zeevaart, *J. Plant Growth Regul.* 9:237-242 (1990)).

10 The expression of a gene in a plant is directed such that the gene has the same temporal and spatial expression pattern of GA4H. The gene is operably linked to the regulatory sequences of GA4H DNA to create an expression module, and a plant is then transformed with the expression module. One can examine the pattern of expression of the endogenous *GA4H* gene using a promoter-glucuronidase (GUS) gene fusion. The data from this analysis is used  
15 to design plant organ-specific promoters and cDNA gene fusions in order to manipulate the GA biosynthesis in specific plant organs.

#### *Immunoblot Analyses*

Proteins from 2-week-old *Arabidopsis* seedlings are extracted and fractionated by centrifugation at 10,000 g for 10 min and then at 100,000 g for 90  
20 min at 4°C (Bensen and Zeevaart, *J. Plant Growth Regul.* 9:237-242, 1990). The 100,000 g supernatant fractions (50 mg each) are loaded on an 8% SDS-PAGE gel, electrophoresed and transferred to a GeneScreen membrane (Du Pont-New England Nuclear). Immunoblot analysis is carried out as described (Sambrook  
25 *et al. Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. 1989). The membrane is incubated with a GA4H antisera (primary antibody), then with 2500-fold diluted peroxidase-conjugated goat anti-rabbit antisera (secondary antibody, Sigma), and detected using the enhanced chemiluminescence reagent (ECL, Amersham) followed by autoradiography.

**EXAMPLE 8*****Over-Expression of GA4H Proteins in E. coli and the Procedure for Generating GA4H Antibodies***

Methods for heterologous expression of DNA clones in *E. coli* are known in the art (Chiang *et al.*, *Plant Cell* 7: 195-201 (1995), Phillips *et al.*, 108:1049-1057 (1995), Wu *et al.*, *Plant Physiol.* 110:547-554 (1996), Yamaguchi *et al.*, *Plant J.* 10:203-213 (1996)). Plasmids containing DNA encoding a GA4H protein are transformed into DE3 lysogenic *E. coli* strain BL21(DE3) (Studier *et al.*, *Methods Enzymol.* 185:60-89 (1990)). The expression of the GA4H cDNA is induced by the addition of 0.4 mM isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) at absorbance (600 nm)=0.8 with 2 hour incubation at 37°C. Thirty ml of cell cultures are harvested by centrifugation, washed and resuspended in 10 ml of 50 mM Tris (pH 8.0), 2 mM EDTA. The cells are sonicated on ice with a Branson microtip at a setting of 4, with four 20-sec pulses. The sonicate is mixed with 1% Triton X-100, incubated on ice for 5 min and then centrifuged at 12000 g for 10 min at 4°C to isolate inclusion bodies (Marston, *DNA Cloning: A Practical Approach*, Oxford England: IRL Press, 1987, with slight modification).

Alternatively, full-length cDNA clones may be expressed as fusion proteins similar to Phillips *et al.* (*Plant Physiol.* 108:1049-1057, 1995) by using for example, an Invitrogen (San Diego, CA) Xpress Kit.

The GA4H proteins are purified from the inclusion body fraction of *E. coli* extracts by SDS-polyacrylamide gel electrophoresis, and electroelution with the Electro-separation system (Schleicher & Schuell). Other methods routinely used by those of skill in the art protein purification can also be used. The purified proteins are detected as single bands on SDS-polyacrylamide gels by Coomassie Blue staining. Rabbit antibodies to GA4H proteins are obtained by subcutaneous injection of gel-purified proteins in complete Freund's adjuvant (Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor, NY, Cold Spring Harbor Laboratory, 1988). For N-group analysis, proteins are fractionated by

SDS-polyacrylamide gel electrophoresis and then transferred to Immobilon membrane (Millipore) in Tris-Glycine and 10% methanol. The membrane is first stained with Ponceau S, destained in deionized water and the appropriate protein bands excised for N-group analysis.

5           The antibodies obtained should be useful for identifying cells or tissues expressing GA4H. A method to accomplish this objective comprises the steps of: a) incubating said cells or said tissues with an agent capable of binding to the GA4H protein or the RNA encoding GA4H; and b) detecting the presence of the bound agent.

10

***Example 9***  
***Modulating the Translation of RNA Encoding GA4H Protein***

The translation of RNA encoding GA4H protein in a plant is modulated by generating an expression vector encoding antisense GA4 HRNA. The plant is then transfected with the expression vector encoding the antisense GA4H RNA.

15

***Example 10***  
***Cloning DNA Encoding GA4H Protein***

20

A DNA molecule encoding the GA4H protein is cloned by hybridizing a desired DNA molecule to the sequences or antisense sequences of for example, DNA SEQ ID No. 5 or DNA SEQ ID No.6 under stringent hybridization conditions. Those DNA molecules hybridizing to the probe sequences are selected and transformed into a host cell. The transformants that express GA4H are selected and cloned.

One possible set of hybridization conditions for the cloning of the DNA encoding GA4H protein is as follows:

25

- 1) prehybridizing for 1 hour;
- 2) hybridizing overnight at 65°C in the hybridization buffer; and
- 3) washing once for 15 minutes in 2xSSC at room temperature, then two times for 30 minutes in 0.1xSSC and 0.1% SDS at 60°C.

**Example 11**  
**Stimulating Plant Stem Elongation**

Plant stem elongation is stimulated by inserting a DNA construct encoding the amino acid sequence of a GA4H protein into a transgenic plant. The transgenic plant is produced by any of several methods known in the art including those previously described in this specification.

The stem elongation may be stimulated in *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersicon*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hererocallis*, *Nemesia*, *Pelargonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Sencia*, *Salpiglossis*, *Cucumis*, *Browalia*, *Glycine*, *Lolium*, *Zea*, *Triticum*, *Sorghum*, *Malus*, *Apium*, and *Datura*.

**Example 12**  
**Producing Dwarf Plants**

Dwarf plants are produced by blocking the *GA4H* gene by homologous recombination, or by transforming with a GA4H anti-sense DNA in order to produce transgenic plants. A cDNA sequence can be used to construct the antisense construct which is then transformed into a plant by using an *Agrobacterium* vector (Zhang *et al.*, *Plant Cell* 4: 1575-1588 (Dec. 1992)). Even partial antisense sequences can be used as antisense and can interfere with the cognate endogenous genes (van der Krol *et al.*, *Plant Mol. Biol.* 14: 457-466 (1990)). The plant is transformed with the antisense construct according to the protocol of Valvekens *et al.*, *Proc. Natl. Acad. Sci, USA* 85:5536-5540 (1988).

Dwarf plants are known to be commercially valuable. For example, dwarf trees for apples, cherries, peaches, pears and nectarines are commercially available (Burpee Gardens Catalogue 1994, pages 122-123).



**Example 13**  
**Molecular Weight Markers**

The GA4H1 and GA4H2 proteins produced recombinantly are purified by routine methods in the art (*Current Protocol in Molecular Biology*, Vol. 2, Chap. 10, John Wiley & Sons, Publishers (1994)). Because the deduced amino acid sequence is known, the molecular weight of these proteins can be precisely determined, and the proteins can be used as molecular weight markers for gel electrophoresis. The calculated molecular weights of the GA4H1 and GA4H2 proteins based on the deduced amino acid sequences are 39086 daltons and 38740 daltons respectively.

**Conclusions**

A genomic clone, comprising the sequences encoding the GA4H1 and GA4H2 proteins was obtained. The GA4H1 and GA4H2 proteins are homologues of the GA4 protein. It is believed that the GA4 locus encodes an hydroxylase involved in gibberellin biosynthesis.

All references mentioned herein are fully incorporated by reference into the disclosure.

Having now fully described the invention by way of illustration and example for purposes of clarity and understanding, it will be apparent to those of ordinary skill in the art that certain changes and modifications may be made in the disclosed embodiments, and such modifications are intended to be within the scope of the present invention.

-52-

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: THE GENERAL HOSPITAL CORPORATION  
FRUIT STREET  
BOSTON, MA 02114  
UNITED STATES OF AMERICA
- APPLICANT/INVENTOR: GOODMAN, HOWARD M.  
NGUYEN, LONG V.  
CHIANG, HUI-HWA
- (ii) TITLE OF INVENTION: GA4 HOMOLOGUE DNA, PROTEIN AND METHODS  
OF USE
- (iii) NUMBER OF SEQUENCES: 29
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
  - (B) STREET: 1100 NEW YORK AVE., SUITE 600
  - (C) CITY: WASHINGTON
  - (D) STATE: DC
  - (E) COUNTRY: USA
  - (F) ZIP: 20005
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: To be assigned
  - (B) FILING DATE: Herewith
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 60/050,615
  - (B) FILING DATE: 24-JUN-1997
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: CIMBALA, MICHELE A.
  - (B) REGISTRATION NUMBER: 33,851
  - (C) REFERENCE/DOCKET NUMBER: 0609.439PC01/MAC/LBB
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: (202)371-2600
  - (B) TELEFAX: (202)371-2540

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1228 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: both
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA

-53-

## (ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 67..1140

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATAAGAAAAA AAACACAAAC ATCTATCAAA TTTACAAAGT TTTAAAACTA ATTAAAAAAG	60
AGCAAG ATG CCT GCT ATG TTA ACA GAT GTG TTT AGA GGC CAT CCC ATT	108
Met Pro Ala Met Leu Thr Asp Val Phe Arg Gly His Pro Ile	
1 5 10	
CAC CTC CCA CAC TCT CAC ATA CCT GAC TTC ACA TCT CTC CGG GAG CTC	156
His Leu Pro His Ser His Ile Pro Asp Phe Thr Ser Leu Arg Glu Leu	
15 20 25 30	
CCG GAT TCT TAC AAG TGG ACC CCT AAA GAC GAT CTC CTC TTC TCC GCT	204
Pro Asp Ser Tyr Lys Trp Thr Pro Lys Asp Asp Leu Leu Phe Ser Ala	
35 40 45	
GCT CCT TCT CCT CCG GCC ACC GGT GAA AAC ATC CCT CTC ATC GAC CTC	252
Ala Pro Ser Pro Pro Ala Thr Gly Glu Asn Ile Pro Leu Ile Asp Leu	
50 55 60	
GAC CAC CCG GAC GCG ACT AAC CAA ATC GGT CAT GCA TGT AGA ACT TGG	300
Asp His Pro Asp Ala Thr Asn Gln Ile Gly His Ala Cys Arg Thr Trp	
65 70 75	
GGT GCC TTC CAA ATC TCA AAC CAC GGC GTG CCT TTG GGA CTT CTC CAA	348
Gly Ala Phe Gln Ile Ser Asn His Gly Val Pro Leu Gly Leu Leu Gln	
80 85 90	
GAC ATT GAG TTT CTC ACC GGT AGT CTC TTC GGG CTA CCT GTC CAA CGC	396
Asp Ile Glu Phe Leu Thr Gly Ser Leu Phe Gly Leu Pro Val Gln Arg	
95 100 105 110	
AAG CTT AAG TCT GCT CGG TCG GAG ACA GGT GTG TCC GGC TAC GGC GTC	444
Lys Leu Lys Ser Ala Arg Ser Glu Thr Gly Val Ser Gly Tyr Gly Val	
115 120 125	
GCT CGT ATC GCA TCT TTC TTC AAT AAG CAA ATG TGG TCC GAA GGT TTC	492
Ala Arg Ile Ala Ser Phe Phe Asn Lys Gln Met Trp Ser Glu Gly Phe	
130 135 140	
ACC ATC ACT GGC TCG CCT CTC AAC GAT TTC CGT AAA CTT TGG CCC CAA	540
Thr Ile Thr Gly Ser Pro Leu Asn Asp Phe Arg Lys Leu Trp Pro Gln	
145 150 155	
CAT CAC CTC AAC TAC TGC GAT ATC GTT GAA GAG TAC GAG GAA CAT ATG	588
His His Leu Asn Tyr Cys Asp Ile Val Glu Glu Tyr Glu Glu His Met	
160 165 170	
AAA AAG TTG GCA TCG AAA TTG ATG TGG TTA GCA CTA AAT TCA CTT GGG	636
Lys Lys Leu Ala Ser Lys Leu Met Trp Leu Ala Leu Asn Ser Leu Gly	
175 180 185 190	
GTC AGC GAA GAA GAC ATT GAA TGG GCC AGT CTC AGT TCA GAT TTA AAC	684
Val Ser Glu Glu Asp Ile Glu Trp Ala Ser Leu Ser Ser Asp Leu Asn	
195 200 205	
TGG GCC CAA GCT GCT CTC CAG CTA AAT CAC TAC CCG GTT TGT CCT GAA	732

-54-

Trp	Ala	Gln	Ala	Ala	Leu	Gln	Leu	Asn	His	Tyr	Pro	Val	Cys	Pro	Glu		
			210					215					220				
CCG	GAC	CGA	GCC	ATG	GGT	CTA	GCA	GCT	CAT	ACC	GAC	TCC	ACC	CTC	CTA		780
Pro	Asp	Arg	Ala	Met	Gly	Leu	Ala	Ala	His	Thr	Asp	Ser	Thr	Leu	Leu		
		225					230					235					
ACC	ATT	CTG	TAC	CAG	AAC	AAT	ACC	GCC	GGT	CTA	CAA	GTA	TTT	CGC	GAT		828
Thr	Ile	Leu	Tyr	Gln	Asn	Asn	Thr	Ala	Gly	Leu	Gln	Val	Phe	Arg	Asp		
	240					245					250						
GAT	CTT	GGT	TGG	GTC	ACC	GTG	CCA	CCG	TTT	CCT	GGC	TCG	CTC	GTG	GTT		876
Asp	Leu	Gly	Trp	Val	Thr	Val	Pro	Pro	Phe	Pro	Gly	Ser	Leu	Val	Val		
255					260					265					270		
AAC	GTT	GGT	GAC	CTC	TTC	CAC	ATC	CTA	TCC	AAT	GGA	TTG	TTT	AAA	AGC		924
Asn	Val	Gly	Asp	Leu	Phe	His	Ile	Leu	Ser	Asn	Gly	Leu	Phe	Lys	Ser		
				275					280					285			
GTG	TTG	CAC	CGC	GCT	CGG	GTT	AAC	CAA	ACC	AGA	GCC	CGG	TTA	TCT	GTA		972
Val	Leu	His	Arg	Ala	Arg	Val	Asn	Gln	Thr	Arg	Ala	Arg	Leu	Ser	Val		
			290					295					300				
GCA	TTC	CTT	TGG	GGT	CCG	CAA	TCT	GAT	ATC	AAG	ATA	TCA	CCT	GTA	CCG		1020
Ala	Phe	Leu	Trp	Gly	Pro	Gln	Ser	Asp	Ile	Lys	Ile	Ser	Pro	Val	Pro		
		305					310					315					
AAG	CTG	GTT	AGT	CCC	GTT	GAA	TCG	CCT	CTA	TAC	CAA	TCG	GTG	ACA	TGG		1068
Lys	Leu	Val	Ser	Pro	Val	Glu	Ser	Pro	Leu	Tyr	Gln	Ser	Val	Thr	Trp		
	320					325					330						
AAA	GAG	TAT	CTT	CGA	ACA	AAA	GCA	ACT	CAC	TTC	AAC	AAA	GCT	CTT	TCA		1116
Lys	Glu	Tyr	Leu	Arg	Thr	Lys	Ala	Thr	His	Phe	Asn	Lys	Ala	Leu	Ser		
335					340				345					350			
ATG	ATT	AGA	AAT	CAC	AGA	GAA	GAA	TGATTAGATA	ATAATAGTTG	TGATCTACTA							1170
Met	Ile	Arg	Asn	His	Arg	Glu	Glu										
				355													
GTTAGTTTGA	TTAATAAATT	GTTGTAAATG	ATTCAGCAA	TATGATTTGT	TTGTCCTC												1228

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 358 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Pro	Ala	Met	Leu	Thr	Asp	Val	Phe	Arg	Gly	His	Pro	Ile	His	Leu		
1				5					10					15			
Pro	His	Ser	His	Ile	Pro	Asp	Phe	Thr	Ser	Leu	Arg	Glu	Leu	Pro	Asp		
			20					25					30				
Ser	Tyr	Lys	Trp	Thr	Pro	Lys	Asp	Asp	Leu	Leu	Phe	Ser	Ala	Ala	Pro		
		35					40					45					

-55-

Ser Pro Pro Ala Thr Gly Glu Asn Ile Pro Leu Ile Asp Leu Asp His  
 50 55 60  
 Pro Asp Ala Thr Asn Gln Ile Gly His Ala Cys Arg Thr Trp Gly Ala  
 65 70 75 80  
 Phe Gln Ile Ser Asn His Gly Val Pro Leu Gly Leu Leu Gln Asp Ile  
 85 90 95  
 Glu Phe Leu Thr Gly Ser Leu Phe Gly Leu Pro Val Gln Arg Lys Leu  
 100 105 110  
 Lys Ser Ala Arg Ser Glu Thr Gly Val Ser Gly Tyr Gly Val Ala Arg  
 115 120 125  
 Ile Ala Ser Phe Phe Asn Lys Gln Met Trp Ser Glu Gly Phe Thr Ile  
 130 135 140  
 Thr Gly Ser Pro Leu Asn Asp Phe Arg Lys Leu Trp Pro Gln His His  
 145 150 155 160  
 Leu Asn Tyr Cys Asp Ile Val Glu Glu Tyr Glu Glu His Met Lys Lys  
 165 170 175  
 Leu Ala Ser Lys Leu Met Trp Leu Ala Leu Asn Ser Leu Gly Val Ser  
 180 185 190  
 Glu Glu Asp Ile Glu Trp Ala Ser Leu Ser Ser Asp Leu Asn Trp Ala  
 195 200 205  
 Gln Ala Ala Leu Gln Leu Asn His Tyr Pro Val Cys Pro Glu Pro Asp  
 210 215 220  
 Arg Ala Met Gly Leu Ala Ala His Thr Asp Ser Thr Leu Leu Thr Ile  
 225 230 235 240  
 Leu Tyr Gln Asn Asn Thr Ala Gly Leu Gln Val Phe Arg Asp Asp Leu  
 245 250 255  
 Gly Trp Val Thr Val Pro Pro Phe Pro Gly Ser Leu Val Val Asn Val  
 260 265 270  
 Gly Asp Leu Phe His Ile Leu Ser Asn Gly Leu Phe Lys Ser Val Leu  
 275 280 285  
 His Arg Ala Arg Val Asn Gln Thr Arg Ala Arg Leu Ser Val Ala Phe  
 290 295 300  
 Leu Trp Gly Pro Gln Ser Asp Ile Lys Ile Ser Pro Val Pro Lys Leu  
 305 310 315 320  
 Val Ser Pro Val Glu Ser Pro Leu Tyr Gln Ser Val Thr Trp Lys Glu  
 325 330 335  
 Tyr Leu Arg Thr Lys Ala Thr His Phe Asn Lys Ala Leu Ser Met Ile  
 340 345 350  
 Arg Asn His Arg Glu Glu  
 355

(2) INFORMATION FOR SEQ ID NO:3:

-56-

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 159 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: both  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
CTTAAGTCTG CTCGGTCGGA GACAGGTGTG TCCGGCTACG GCGTCGCTCG TATCGCATCT      60
TTCTTCAATA AGCAAATGTG GTCCGAAGGT TTCACCATCA CTGGCTCGCC TCTCAACGAT      120
TTCCGTAAAC TTTGGCCCCA ACATCACCTC AACTACTGC                               159
```

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 140 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: both  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
GTGGTTAGCA CTAAATTCAC TTGGGGTCAG CGAAGAAGAC ATTGAATGGG CCAGTCTCAG      60
TTCAGATTTA AACTGGGCCC AAGCTGCTCT CCAGCTAAAT CACTACCCGG TTTGTCCTGA      120
ACCGGACCGA GCCATGGGTC                               140
```

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 1133 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: both  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
GTTTATGTGA TGAGCATCCC ATTCTCTCAT TAGTTCACAA GTCATGCCTT CACTAGCAGA      60
AGAGATATGT ATTGGTAACT TAGGCAGTCT CCAAACACTC CCCGAGTCGT TCACCTGGAA      120
ACTCACAGCC GCCGACTCCC TTCTGCGTCC CTCCTCCGCC GTCTCATTCG ACGCAGTGGA      180
AGAGTCCATT CCTGTGATCG ACCTCTCTAA TCCTGACGTT ACCACCCTCA TTGGAGATGC      240
```

-57-

CTCCAAAACA TGGGGAGCGT TTCAGATAGC CAACCACGGG ATTTCTCAGA AGCTTCTCGA	300
TGATATCGAG TCTCTGTCCA AAACCTATT CGACATGCCG TCAGAGAGGA AGCTTGAAGC	360
GGCTTCCTCC GATAAAGGAG TTAGTGGCTA CGGAGAACCT CGAATCTCCC CCTTTTTCGA	420
GAAGAAAATG TGGTCTGAAG GGTTTACTAT TGCCGATGAC TCCTACCGCA ACCATTTCAA	480
TACTCTTTGG CCTCATGATC ACACCAAGTA CTGCGGTATA ATCCAAGAAT ACGTGGACGA	540
AATGGAAAAA TTAGCAAGCA GACTTCTGTA TTGCACATTA GGCTCACTTG GTGTCACCGT	600
GGAAGACATT GAATGGGCTC ACAAGCTAGA GAAATCTGGA TCAAAAGTGG GCAGAGGCGC	660
CATACGACTA AACCCTACC CGGTTTGTCC TGAACCAGAA CGAGCCATGG GTCTAGCCGC	720
TCATACAGAC TCCACTATCC TAACCATTCT GCACCAGAGC AACACGGGAG GGCTACAAGT	780
GTTCAGGGAA GAGTCCGGTT GGGTCACGGT TGAGCCGGCT CCTGGTGTCC TCGTGGTCAA	840
CATGGGTGAT CTCTTTCACA TCTTATCGAA CGGGAAAATC CCAAGCGTGG TTCATCGAGC	900
CAAAGTTAAC CATACTCGGT CAAGAATTTT GATTGCGTAC TTATGGGGTG GTCCAGCTGG	960
TGATGTGCAA ATCGCACCTA TCTCTAAGTT AACCGGTCCG GCTGAACCGT CTCTTTACCG	1020
GTCAATTACA TGGAAAGAGT ATCTCCAAAT AAAGTATGGG GTTTTCGACA AGGCCATGGA	1080
CGCAATTAGG GTCGTTAATC CCACCAATTA AATCTCCTTC TCAAATACTC TCT	1133

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1610 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 86..556

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 966..1559

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

ACATATGTGT GTAGTATCTA TGCATATATA TCCAAAGTAA TTGTTTATGT GATGAGCATC	60
CCATTCTCTC ATTAGTTCAC AAGTC ATG CCT TCA CTA GCA GAA GAG ATA TGT	112
Met Pro Ser Leu Ala Glu Glu Ile Cys	
1 5	
ATT GGT AAC TTA GGC AGT CTC CAA ACA CTC CCC GAG TCG TTC ACC TGG	160
Ile Gly Asn Leu Gly Ser Leu Gln Thr Leu Pro Glu Ser Phe Thr Trp	
10 15 20 25	

-58-

AAA CTC ACA GCC GCC GAC TCC CTT CTG CGT CCC TCC TCC GCC GTC TCA	208
Lys Leu Thr Ala Ala Asp Ser Leu Leu Arg Pro Ser Ser Ala Val Ser	
30 35 40	
TTC GAC GCA GTG GAA GAG TCC ATT CCT GTG ATC GAC CTC TCT AAT CCT	256
Phe Asp Ala Val Glu Glu Ser Ile Pro Val Ile Asp Leu Ser Asn Pro	
45 50 55	
GAC GTT ACC ACC CTC ATT GGA GAT GCC TCC AAA ACA TGG GGA GCG TTT	304
Asp Val Thr Thr Leu Ile Gly Asp Ala Ser Lys Thr Trp Gly Ala Phe	
60 65 70	
CAG ATA GCC AAC CAC GGG ATT TCT CAG AAG CTT CTC GAT GAT ATC GAG	352
Gln Ile Ala Asn His Gly Ile Ser Gln Lys Leu Leu Asp Asp Ile Glu	
75 80 85	
TCT CTG TCC AAA ACC CTA TTC GAC ATG CCG TCA GAG AGG AAG CTT GAA	400
Ser Leu Ser Lys Thr Leu Phe Asp Met Pro Ser Glu Arg Lys Leu Glu	
90 95 100 105	
GCG GCT TCC TCC GAT AAA GGA GTT AGT GGC TAC GGA GAA CCT CGA ATC	448
Ala Ala Ser Ser Asp Lys Gly Val Ser Gly Tyr Gly Glu Pro Arg Ile	
110 115 120	
TCC CCC TTT TTC GAG AAG AAA ATG TGG TCT GAA GGG TTT ACT ATT GCC	496
Ser Pro Phe Phe Glu Lys Lys Met Trp Ser Glu Gly Phe Thr Ile Ala	
125 130 135	
GAT GAC TCC TAC CGC AAC CAT TTC AAT ACT CTT TGG CCT CAT GAT CAC	544
Asp Asp Ser Tyr Arg Asn His Phe Asn Thr Leu Trp Pro His Asp His	
140 145 150	
ACC AAG TAC TGG TAACGTCTAT TACACACACA TATATATATT TTTTGCTTAT	596
Thr Lys Tyr Trp	
155	
TTCGCAAAAG TGTGGCAAAG GAAATTGCAC ACTTTTTTTTT TGCAC TAAGA CTTAGTTATT	656
ATTAAAAGTG TTAAATGTT TTTTCTGTG CATAAAAAAG TGTTTATATG TTCCGAGTAA	716
TTGATGTTTA TGATTAGTGA TAACTGATAA CACATAGAGT GTAGCCTTCA AAGTTTCTAA	776
TTAAATAGTT TGAGCAACAT CCTTATATTT TATGAAGTAG TACTTCTTAT TGCATATTAC	836
AGCAAATTAA AGTACCAAAG TCTCTATGAA ATGTGATAAT TTGGCTAATG TCGAGGTCTT	896
AACATTAGAT TACCAAAAAC CTTAATTACT GTAAATTGTA TTTGCTTTTC ATTTTGGTA	956
TTGTGCAGC GGT ATA ATC CAA GAA TAC GTG GAC GAA ATG GAA AAA TTA	1004
Gly Ile Ile Gln Glu Tyr Val Asp Glu Met Glu Lys Leu	
1 5 10	
GCA AGC AGA CTT CTG TAT TGC ACA TTA GGC TCA CTT GGT GTC ACC GTG	1052
Ala Ser Arg Leu Leu Tyr Cys Thr Leu Gly Ser Leu Gly Val Thr Val	
15 20 25	
GAA GAC ATT GAA TGG GCT CAC AAG CTA GAG AAA TCT GGA TCA AAA GTG	1100
Glu Asp Ile Glu Trp Ala His Lys Leu Glu Lys Ser Gly Ser Lys Val	
30 35 40 45	
GGC AGA GGC GCC ATA CGA CTA AAC CAC TAC CCG GTT TGT CCT GAA CCA	1148
Gly Arg Gly Ala Ile Arg Leu Asn His Tyr Pro Val Cys Pro Glu Pro	



-59-

50										55					60					
GAA	CGA	GCC	ATG	GGT	CTA	GCC	GCT	CAT	ACA	GAC	TCC	ACT	ATC	CTA	ACC	1196				
Glu	Arg	Ala	Met	Gly	Leu	Ala	Ala	His	Thr	Asp	Ser	Thr	Ile	Leu	Thr					
			65					70					75							
ATT	CTG	CAC	CAG	AGC	AAC	ACG	GGA	GGG	CTA	CAA	GTG	TTC	AGG	GAA	GAG	1244				
Ile	Leu	His	Gln	Ser	Asn	Thr	Gly	Gly	Leu	Gln	Val	Phe	Arg	Glu	Glu					
		80					85					90								
TCC	GGT	TGG	GTC	ACG	GTT	GAG	CCG	GCT	CCT	GGT	GTC	CTC	GTG	GTC	AAC	1292				
Ser	Gly	Trp	Val	Thr	Val	Glu	Pro	Ala	Pro	Gly	Val	Leu	Val	Val	Asn					
	95					100					105									
ATG	GGT	GAT	CTC	TTT	CAC	ATC	TTA	TCG	AAC	GGG	AAA	ATC	CCA	AGC	GTG	1340				
Met	Gly	Asp	Leu	Phe	His	Ile	Leu	Ser	Asn	Gly	Lys	Ile	Pro	Ser	Val					
110					115				120					125						
GTT	CAT	CGA	GCC	AAA	GTT	AAC	CAT	ACT	CGG	TCA	AGA	ATT	TCG	ATT	GCG	1388				
Val	His	Arg	Ala	Lys	Val	Asn	His	Thr	Arg	Ser	Arg	Ile	Ser	Ile	Ala					
				130					135					140						
TAC	TTA	TGG	GGT	GGT	CCA	GCT	GGT	GAT	GTG	CAA	ATC	GCA	CCT	ATC	TCT	1436				
Tyr	Leu	Trp	Gly	Gly	Pro	Ala	Gly	Asp	Val	Gln	Ile	Ala	Pro	Ile	Ser					
			145					150					155							
AAG	TTA	ACC	GGT	CCG	GCT	GAA	CCG	TCT	CTT	TAC	CGG	TCA	ATT	ACA	TGG	1484				
Lys	Leu	Thr	Gly	Pro	Ala	Glu	Pro	Ser	Leu	Tyr	Arg	Ser	Ile	Thr	Trp					
		160					165					170								
AAA	GAG	TAT	CTC	CAA	ATA	AAG	TAT	GAG	GTT	TTC	GAC	AAG	GCC	ATG	GAC	1532				
Lys	Glu	Tyr	Leu	Gln	Ile	Lys	Tyr	Glu	Val	Phe	Asp	Lys	Ala	Met	Asp					
	175					180					185									
GCA	ATT	AGG	GTC	GTT	AAT	CCC	ACC	AAT	TAAATCTCCT	TCTCAAATAC						1579				
Ala	Ile	Arg	Val	Val	Asn	Pro	Thr	Asn												
190					195															
TCTCTTAATG	AAAAACCTAA	ATTAAATGCG	A													1610				

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 157 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met	Pro	Ser	Leu	Ala	Glu	Glu	Ile	Cys	Ile	Gly	Asn	Leu	Gly	Ser	Leu
1				5					10					15	
Gln	Thr	Leu	Pro	Glu	Ser	Phe	Thr	Trp	Lys	Leu	Thr	Ala	Ala	Asp	Ser
		20						25					30		
Leu	Leu	Arg	Pro	Ser	Ser	Ala	Val	Ser	Phe	Asp	Ala	Val	Glu	Glu	Ser
		35					40					45			
Ile	Pro	Val	Ile	Asp	Leu	Ser	Asn	Pro	Asp	Val	Thr	Thr	Leu	Ile	Gly

-60-

50	55	60
Asp Ala Ser Lys Thr Trp Gly Ala Phe Gln Ile Ala Asn His Gly Ile		
65	70	75
Ser Gln Lys Leu Leu Asp Asp Ile Glu Ser Leu Ser Lys Thr Leu Phe		
	85	90
Asp Met Pro Ser Glu Arg Lys Leu Glu Ala Ala Ser Ser Asp Lys Gly		
	100	105
Val Ser Gly Tyr Gly Glu Pro Arg Ile Ser Pro Phe Phe Glu Lys Lys		
	115	120
Met Trp Ser Glu Gly Phe Thr Ile Ala Asp Asp Ser Tyr Arg Asn His		
	130	135
Phe Asn Thr Leu Trp Pro His Asp His Thr Lys Tyr Trp		
145	150	155

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 198 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Gly Ile Ile Gln Glu Tyr Val Asp Glu Met Glu Lys Leu Ala Ser Arg		
1	5	10
Leu Leu Tyr Cys Thr Leu Gly Ser Leu Gly Val Thr Val Glu Asp Ile		
	20	25
Glu Trp Ala His Lys Leu Glu Lys Ser Gly Ser Lys Val Gly Arg Gly		
	35	40
Ala Ile Arg Leu Asn His Tyr Pro Val Cys Pro Glu Pro Glu Arg Ala		
	50	55
Met Gly Leu Ala Ala His Thr Asp Ser Thr Ile Leu Thr Ile Leu His		
	65	70
Gln Ser Asn Thr Gly Gly Leu Gln Val Phe Arg Glu Glu Ser Gly Trp		
	85	90
Val Thr Val Glu Pro Ala Pro Gly Val Leu Val Val Asn Met Gly Asp		
	100	105
Leu Phe His Ile Leu Ser Asn Gly Lys Ile Pro Ser Val Val His Arg		
	115	120
Ala Lys Val Asn His Thr Arg Ser Arg Ile Ser Ile Ala Tyr Leu Trp		
	130	135
Gly Gly Pro Ala Gly Asp Val Gln Ile Ala Pro Ile Ser Lys Leu Thr		
145	150	155
		160

-61-

Gly Pro Ala Glu Pro Ser Leu Tyr Arg Ser Ile Thr Trp Lys Glu Tyr  
165 170 175

Leu Gln Ile Lys Tyr Glu Val Phe Asp Lys Ala Met Asp Ala Ile Arg  
180 185 190

Val Val Asn Pro Thr Asn  
195

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1105 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: both  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TCATACCAAA	AACATAAAGC	CAAAATATAA	ACACATAAGC	CTTTTAGCAT	GAGTTCAACG	60
TTGAGCGATG	TGTTTAGATC	GCATCCCATT	CACATCCCAC	TCTCAAACCC	ACCTGACTTC	120
AAATCTCTCC	CGGATTCTTA	CACGTGGACT	CCTAAAGATG	ATCTCCTCTT	CTCCGCCTCC	180
GCCTCCGACG	AAACCCTGCC	GCTCATCGAC	CTCTCCGATA	TCCACGTGGC	CACTCTTGTG	240
GGCCATGCTT	GTACCACGTG	GGGAGCGTTC	CAGATCACCA	ACCACGGCGT	CCCCTCGCGA	300
CTTCTCGACG	ACATTGAGTT	CCTCACCGGA	AGTCTTTTCC	GGCTTCCCGT	ACAGCGGAAG	360
CTCAAGGCGG	CTCGGTCAGA	GAATGGCGTC	TCCGGCTACG	GCGTAGCTCG	TATTGCTTCG	420
TTCTTTAATA	AGAAGATGTG	GTCCGAAGGT	TTCACCGTTA	TTGGCTCTCC	CCTCCACGAT	480
TTCCGTAAAC	TCTGGCCCAG	CCACCACCTC	AAATACTGTG	AAATTATTGA	AGAGTATGAA	540
GAACATATGC	AAAAGTTGGC	AGCCAAGTTG	ATGTGGTTCG	CATTAGGTTT	ACTGGGAGTT	600
GAAGAAAAGG	ACATACAATG	GGCCGGGCCCT	AATTCAGACT	TTCAAGGAAC	CCAAGCAGCT	660
ATCCAACATA	ACCATTATCC	AAAATGTCCA	GAACCAGACA	GAGCCATGGG	CCTCGCAGCC	720
CATACAGACT	CGACCCTCAT	GACCATTTCT	TACCAGAACA	ACACCGCCGG	TCTCCAAGTT	780
TTCCGGGATG	ACGTGGGCTG	GGTTACCGCG	CCACCTGTCC	CTGGCTCGCT	GGTGGTCAAC	840
GTCGGTGACT	TGCTCCACAT	TTTAACCAAC	GGAATCTTCC	CGAGCGTGCT	TCACCGAGCC	900
AGGGTTAACC	ACGTCCGATC	TCGGTTCTCA	ATGGCTTACC	TGTGGGGTCC	ACCATCCGAT	960
GTAATGATCT	CTCCACTTCC	CAAACGGTTT	GATCCTCTCC	AATCTCCTCT	CTACCCATCT	1020
CTCACTTGGA	AACAATACCT	TGCTACCAAA	GCTACTCATT	TTAATCAATC	TCTTTCCATT	1080
ATTAGAAATT	AACTGTCTTC	CGACT				1105

-62-

## (2) INFORMATION FOR SEQ ID NO:10:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1690 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 95..565

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 986..1555

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TCACCGATCT ATAAATACAC TCCTCTTCTC CACCAAAAGT ATCATATCAT ACCAAAAACA	60
TAAAGCCAAA ATATAAACAC ATAAGCCTTT TAGC ATG AGT TCA ACG TTG AGC	112
Met Ser Ser Thr Leu Ser	
1 5	
GAT GTG TTT AGA TCG CAT CCC ATT CAC ATC CCA CTC TCA AAC CCA CCT	160
Asp Val Phe Arg Ser His Pro Ile His Ile Pro Leu Ser Asn Pro Pro	
10 15 20	
GAC TTC AAA TCT CTC CCG GAT TCT TAC ACG TGG ACT CCT AAA GAT GAT	208
Asp Phe Lys Ser Leu Pro Asp Ser Tyr Thr Trp Thr Pro Lys Asp Asp	
25 30 35	
CTC CTC TTC TCC GCC TCC GCC TCC GAC GAA ACC CTG CCG CTC ATC GAC	256
Leu Leu Phe Ser Ala Ser Ala Ser Asp Glu Thr Leu Pro Leu Ile Asp	
40 45 50	
CTC TCC GAT ATC CAC GTG GCC ACT CTT GTG GGC CAT GCT TGT ACC ACG	304
Leu Ser Asp Ile His Val Ala Thr Leu Val Gly His Ala Cys Thr Thr	
55 60 65 70	
TGG GGA GCG TTC CAG ATC ACC AAC CAC GGC GTC CCC TCG CGA CTT CTC	352
Trp Gly Ala Phe Gln Ile Thr Asn His Gly Val Pro Ser Arg Leu Leu	
75 80 85	
GAC GAC ATT GAG TTC CTC ACC GGA AGT CTT TTC CGG CTT CCC GTA CAG	400
Asp Asp Ile Glu Phe Leu Thr Gly Ser Leu Phe Arg Leu Pro Val Gln	
90 95 100	
CGG AAG CTC AAG GCG GCT CGG TCA GAG AAT GGC GTC TCC GGC TAC GGC	448
Arg Lys Leu Lys Ala Ala Arg Ser Glu Asn Gly Val Ser Gly Tyr Gly	
105 110 115	
GTA GCT CGT ATT GCT TCG TTC TTT AAT AAG AAG ATG TGG TCC GAA GGT	496
Val Ala Arg Ile Ala Ser Phe Phe Asn Lys Lys Met Trp Ser Glu Gly	
120 125 130	
TTC ACC GTT ATT GGC TCT CCC CTC CAC GAT TTC CGT AAA CTC TGG CCC	544
Phe Thr Val Ile Gly Ser Pro Leu His Asp Phe Arg Lys Leu Trp Pro	

-63-

135	140	145	150	
AGC CAC CAC CTC AAA TAC TGG TATCTTTTTC AATGGTTCAT TTTATCAACG				595
Ser His His Leu Lys Tyr Trp				
155				
TTAAGACCAT ATTAACGTAA CGTAACTTAT CTTTGTATGA AAAAAAAAAA AAAAAGGTG				655
GACGTTAGTA CAGTTGACTA TTCAATTGAT ATAGATTCGG GAATAATACG AAAAGGGTAA				715
AGTAGAAACC ATTTTTTGCC ATGTCGTAGT TAGTAAAAAG CACAATGAAA ACTCATGGAC				775
CCACCAAAAA GATTACATGA TATAATATAT ATATATATAT TTATATAAAT ATTATATAAT				835
ATATTTATAT AATATTATGT GCAAAAATTA AATGAAAATA AATATTATCA GGAGAATGTG				895
AAATACAGTA TAAGATTTTC CTTTGGCTAC ATGACGATTT CTATAGATTT GAAGGTAAAG				955
ATACTAATTT CATATTATCG ATTCAACAGT GAA ATT ATT GAA GAG TAT GAA GAA				1009
Glu Ile Ile Glu Glu Tyr Glu Glu				
1			5	
CAT ATG CAA AAG TTG GCA GCC AAG TTG ATG TGG TTC GCA TTA GGT TCA				1057
His Met Gln Lys Leu Ala Ala Lys Leu Met Trp Phe Ala Leu Gly Ser				
10		15	20	
CTG GGA GTT GAA GAA AAG GAC ATA CAA TGG GCC GGG CCT AAT TCA GAC				1105
Leu Gly Val Glu Glu Lys Asp Ile Gln Trp Ala Gly Pro Asn Ser Asp				
25		30	35	40
TTT CAA GGA ACC CAA GCA GCT ATC CAA CTA AAC CAT TAT CCA AAA TGT				1153
Phe Gln Gly Thr Gln Ala Ala Ile Gln Leu Asn His Tyr Pro Lys Cys				
45		50	55	
CCA GAA CCA GAC AGA GCC ATG GGC CTC GCA GCC CAT ACA GAC TCG ACC				1201
Pro Glu Pro Asp Arg Ala Met Gly Leu Ala Ala His Thr Asp Ser Thr				
60		65	70	
CTC ATG ACC ATT CTG TAC CAG AAC AAC ACC GCC GGT CTC CAA GTT TTC				1249
Leu Met Thr Ile Leu Tyr Gln Asn Asn Thr Ala Gly Leu Gln Val Phe				
75		80	85	
CGG GAT GAC GTG GGC TGG GTT ACC GCG CCA CCT GTC CCT GGC TCG CTG				1297
Arg Asp Asp Val Gly Trp Val Thr Ala Pro Pro Val Pro Gly Ser Leu				
90		95	100	
GTG GTC AAC GTC GGT GAC TTG CTC CAC ATT TTA ACC AAC GGA ATC TTC				1345
Val Val Asn Val Gly Asp Leu Leu His Ile Leu Thr Asn Gly Ile Phe				
105		110	115	120
CCG AGC GTG CTT CAC CGA GCC AGG GTT AAC CAC GTC CGA TCT CGG TTC				1393
Pro Ser Val Leu His Arg Ala Arg Val Asn His Val Arg Ser Arg Phe				
125		130	135	
TCA ATG GCT TAC CTG TGG GGT CCA CCA TCC GAT GTA ATG ATC TCT CCA				1441
Ser Met Ala Tyr Leu Trp Gly Pro Pro Ser Asp Val Met Ile Ser Pro				
140		145	150	
CTT CCC AAA CTG GTT GAT CCT CTC CAA TCT CCT CTC TAC CCA TCT CTC				1489
Leu Pro Lys Leu Val Asp Pro Leu Gln Ser Pro Leu Tyr Pro Ser Leu				
155		160	165	

-64-

ACT TGG AAA CAA TAC CTT GCT ACC AAA GCT ACT CAT TTT AAT CAA TCT 1537  
 Thr Trp Lys Gln Tyr Leu Ala Thr Lys Ala Thr His Phe Asn Gln Ser  
 170 175 180

CTT TCC ATT ATT AGA AAT TAACTGTCTT CCGACTGAAT TTCTTGATTT 1585  
 Leu Ser Ile Ile Arg Asn  
 185 190

TCAGATTTTA CTATTTATTT TCTTAGTAAT ATGATGATAT CTATTACTGT TTCGATTTTA 1645

GATGAGTGGT TCTTCAAATT CACAATTAGT AGCTTAATAT TGATT 1690

## (2) INFORMATION FOR SEQ ID NO:11:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 157 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met Ser Ser Thr Leu Ser Asp Val Phe Arg Ser His Pro Ile His Ile  
 1 5 10 15

Pro Leu Ser Asn Pro Pro Asp Phe Lys Ser Leu Pro Asp Ser Tyr Thr  
 20 25 30

Trp Thr Pro Lys Asp Asp Leu Leu Phe Ser Ala Ser Ala Ser Asp Glu  
 35 40 45

Thr Leu Pro Leu Ile Asp Leu Ser Asp Ile His Val Ala Thr Leu Val  
 50 55 60

Gly His Ala Cys Thr Thr Trp Gly Ala Phe Gln Ile Thr Asn His Gly  
 65 70 75 80

Val Pro Ser Arg Leu Leu Asp Asp Ile Glu Phe Leu Thr Gly Ser Leu  
 85 90 95

Phe Arg Leu Pro Val Gln Arg Lys Leu Lys Ala Ala Arg Ser Glu Asn  
 100 105 110

Gly Val Ser Gly Tyr Gly Val Ala Arg Ile Ala Ser Phe Phe Asn Lys  
 115 120 125

Lys Met Trp Ser Glu Gly Phe Thr Val Ile Gly Ser Pro Leu His Asp  
 130 135 140

Phe Arg Lys Leu Trp Pro Ser His His Leu Lys Tyr Trp  
 145 150 155

## (2) INFORMATION FOR SEQ ID NO:12:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 190 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

-65-

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Glu Ile Ile Glu Glu Tyr Glu Glu His Met Gln Lys Leu Ala Ala Lys  
 1 5 10 15  
 Leu Met Trp Phe Ala Leu Gly Ser Leu Gly Val Glu Glu Lys Asp Ile  
 20 25 30  
 Gln Trp Ala Gly Pro Asn Ser Asp Phe Gln Gly Thr Gln Ala Ala Ile  
 35 40 45  
 Gln Leu Asn His Tyr Pro Lys Cys Pro Glu Pro Asp Arg Ala Met Gly  
 50 55 60  
 Leu Ala Ala His Thr Asp Ser Thr Leu Met Thr Ile Leu Tyr Gln Asn  
 65 70 75 80  
 Asn Thr Ala Gly Leu Gln Val Phe Arg Asp Asp Val Gly Trp Val Thr  
 85 90 95  
 Ala Pro Pro Val Pro Gly Ser Leu Val Val Asn Val Gly Asp Leu Leu  
 100 105 110  
 His Ile Leu Thr Asn Gly Ile Phe Pro Ser Val Leu His Arg Ala Arg  
 115 120 125  
 Val Asn His Val Arg Ser Arg Phe Ser Met Ala Tyr Leu Trp Gly Pro  
 130 135 140  
 Pro Ser Asp Val Met Ile Ser Pro Leu Pro Lys Leu Val Asp Pro Leu  
 145 150 155 160  
 Gln Ser Pro Leu Tyr Pro Ser Leu Thr Trp Lys Gln Tyr Leu Ala Thr  
 165 170 175  
 Lys Ala Thr His Phe Asn Gln Ser Leu Ser Ile Ile Arg Asn  
 180 185 190

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GTGGTTAGCA CTAAATTCAC

20

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid

-66-

- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GACCCATGGC TCGGTCCGGT

20

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 23 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: both
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GCTCTAGAGA GTATTGAGA AGG

23

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: both
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GTTTACTATT GCCGATGACT

20

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: both
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CAATACCAAA AATGAAAAGC

20



-67-

## (2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: both
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CTCCTACCGC AACCATTTC

19

## (2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 28 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: both
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TCCCCCGGGT TTATGTGATG AGCATCCC

28

## (2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: both
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CCAAAGTAAT TGTTTATGTG

20

## (2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: both
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

-68-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

AATTTAGGTT TTTCATTAAG

20

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GTAGTGTTT AGTCGTATGG

20

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

AAACTTGGG GACCGGCGG

19

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

TATCATGTAA TCTTTTGG

19

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both

-69-

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CCGGCTTCCC GTACAGCGG

19

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: both

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

AATCAAGAAA TTCAGTCGG

19

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: both

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GGAATTCATA CCAAAAACAT AAAGCC

26

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: both

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CTAGTTTCTT TCTTCCACG

19

-70-

## (2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: both
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: cDNA

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

TAGCTGCATC TTCTTTACC

19

**INDICATIONS RELATING TO A DEPOSITED MICROORGANISM**  
(PCT Rule 13bis)

<b>A.</b> The indications made below relate to the microorganism referred to in the description on page <u>4</u> , line <u>25-26</u> .	
<b>B. IDENTIFICATION OF DEPOSIT</b> <span style="float: right;">Further deposits are identified on an additional sheet <input type="checkbox"/></span>	
Name of depositary institution  American Type Culture Collection	
Address of depositary institution <i>(including postal code and country)</i>  <div style="display: flex; justify-content: space-between;"> <div>           12301 Parklawn Drive            Rockville, Maryland 20852            United States of America         </div> <div style="text-align: center;">           Now at         </div> <div>           10801 University Boulevard            Manassas, Virginia 20110-2209            United States of America         </div> </div>	
Date of deposit May 20, 1997	Accession Number ATCC 98436
<b>C. ADDITIONAL INDICATIONS</b> <i>(leave blank if not applicable)</i> <span style="float: right;">This information is continued on an additional sheet <input type="checkbox"/></span>	
Arabidopsis thaliana genomic DNA of GA4H1 and GA4H2 genes cloned into pBSKS(+) (Stratagene) vector pLVN103 in DH5α	
<b>D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE</b> <i>(if the indications are not for all designated States)</i>  <div style="height: 50px;"></div>	
<b>E. SEPARATE FURNISHING OF INDICATIONS</b> <i>(leave blank if not applicable)</i>  The indications listed below will be submitted to the international Bureau later <i>(specify the general nature of the indications, e.g., "Accession Number of Deposit")</i>  <div style="height: 100px;"></div>	

<p align="center">For receiving Office use only</p> <div style="border: 1px solid black; padding: 5px;"> <input checked="" type="checkbox"/> This sheet was received with the international application       </div> <div style="border: 1px solid black; padding: 5px;">         Authorized officer  <div style="display: flex; align-items: center;"> <div style="margin-right: 20px;">PAUL F. URRUTIA</div> </div> </div>	<p align="center">For International Bureau use only</p> <div style="border: 1px solid black; padding: 5px;"> <input type="checkbox"/> This sheet was received by the International Bureau on:       </div> <div style="border: 1px solid black; padding: 5px;">         Authorized officer       </div>
--	---

***What Is Claimed Is:***

1. A purified DNA molecule comprising a DNA sequence encoding the amino acid sequence of a GA4 homologue.

5           2. The DNA molecule of claim 1 encoding the amino acid sequence of GA4H1 in Figure 6 (SEQ. ID. No. 7).

3. The DNA molecule of claim 1 encoding the amino acid sequence of GA4H2 in Figure 8 (SEQ. ID. No. 10).

10           4. The DNA molecule of claim 1, wherein said DNA is selected from the group consisting of the genomic DNA's, SEQ ID No. 6 in Figure 6, SEQ ID No. 9 in Figure 8, cDNAs having SEQ ID No. 5 in Figure 5, SEQ. ID. No. 8 in Figure 7 and a degenerate variant of any of said sequences.

5. A DNA molecule comprising a sequence with at least 95% homology to the DNA sequence in any one of claims 1-4.

15           6. A vector comprising the sequence of claim 5.

7. A host transformed with the vector of claim 6.

8. The host of claim 7, wherein said host is selected from the group consisting of bacteria, yeast, plants, insects or mammals.

9. The host of claim 8, wherein said host is a plant cell.

20           10. The host of claim 9, wherein said plant cell is a dicotyledonous plant cell.

11. A plant regenerated from the plant cell of claim 10.
12. Progeny of the plant of claim 11.
13. A propagule of the plant of claim 11.
- 5 14. A seed produced by the progeny of claim 11.
15. Purified GA4H protein.
16. The protein of claim 15, wherein said GA4H protein is an *Arabidopsis* protein.
- 10 17. The GA4H protein of claim 15, wherein said GA4H protein is selected from the group consisting of GA4H1 comprising the amino acid sequence shown in Figure 6 (SEQ. ID NO. 7), GA4H2 comprising the amino acid sequence shown in Figure 8 (SEQ. ID NO. 10) and a functional derivative of said sequences.
- 15 18. The GA4H protein of any one of claims 15-17, wherein said GA4H protein is substantially free of other *A. thaliana* proteins.
19. A cell extract comprising a GA4H protein.
20. The cell extract of claim 21, wherein said GA4H protein is an *Arabidopsis* protein.
- 20 21. The cell extract of claim 26, wherein said GA4H protein comprises the amino acid sequence selected from the group consisting of Figure 6 (SEQ. ID NO. 7) , Figure 8 (SEQ. ID. NO. 10) and a functional derivative of said sequences.

22. The cell extract of any one of claims 19-21 wherein said cell is a prokaryotic cell or a eukaryotic cell.

23. The cell extract of claim 22, wherein said prokaryotic cell is an *E. coli*.

24. The cell extract of claim 22, wherein said eukaryotic cell is a yeast, fungal, insect, mammalian or transgenic plant cell.

25. A cell extract comprising *A. thaliana* GA4H protein, wherein said cell is not *A. thaliana*.

26. A method of making GA4H protein wherein said GA4H protein is substantially free of other *A. thaliana* proteins, said method comprising:

- a) transforming a prokaryotic or eukaryotic cell with a GA4H recombinant expression vector encoding a GA4H protein or a functional derivative of a GA4H protein,
- b) expressing said GA4H protein, and
- c) isolating said GA4H protein substantially free of other *A. thaliana* proteins.

27. The method of claim 26 wherein said GA4H protein is isolated from *E. coli* inclusion bodies.

28. A method of directing the expression of a gene in a plant, such that said gene has the same temporal and spatial expression pattern of a *GA4H*, said method comprising the steps of:

- 1) operably linking said gene to the regulatory sequences of *GA4H* to create an expression module, and
- 2) transforming said plant with said expression module of part (1).



29. A method of modulating the translation of RNA encoding GA4H in a plant comprising the steps of:

- 5                   1) generating an expression vector encoding antisense GA4H RNA;
- 2) transfecting said plant with said expression vector of part (1).

30. An isolated DNA construct wherein said construct consists essentially of a nucleic acid sequence, and wherein said nucleic acid sequence:

- 10                   1) encodes GA4H polypeptide, and
- 2) hybridizes to the sense or antisense sequence of the GA4H DNA when hybridization is performed under stringent hybridization conditions.

31. An isolated DNA molecule encoding a GA4H protein, said DNA molecule prepared by a process comprising:

- 15                   1) hybridizing a desired DNA molecule to the sense or antisense sequence of a GA4H DNA sequence, wherein the hybridization is performed under stringent hybridization conditions;
- 20                   2) selecting those DNA molecules of said population that hybridize to said sequence; and
- 3) selecting DNA molecules of part (2) that encode said GA4H protein.

32. An isolated DNA molecule encoding a GA4H protein as claimed in claims 30 or 31, said DNA molecule prepared by a process comprising:

- 25                   1) prehybridizing for 1 hour;
- 2) hybridizing overnight at 65°C in the hybridization buffer; and

- 3) washing once for 15 minutes in 2xSSC at room temperature, then two times for 30 minutes in 0.1xSSC and 0.1% SDS at 60°C.

5

33. A method of cloning a DNA molecule that encodes a GA4H protein, said method comprising:

- 1) hybridizing a desired DNA molecule to the sense or antisense sequence of GA4H DNA wherein the hybridization is performed under stringent hybridization conditions;
- 2) selecting those DNA molecules of said population that hybridize to said sequence;
- 3) transforming said DNA of part (2) into a host cell; and
- 4) selecting transformants that express said GA4.

10

15

34. The method of claim 33 wherein the hybridization conditions consist essentially of:

- 1) prehybridizing for 1 hour;
- 2) hybridizing overnight at 65°C in the hybridization buffer; and
- 3) washing once for 15 minutes in 2xSSC at room temperature, then two times for 30 minutes in 0.1xSSC and 0.1% SDS at 60°C.

20

35. A method of altering stem elongation, said method comprising inserting a DNA construct encoding the amino acid sequence of a GA4H protein into a transgenic plant.

25

36. A method of producing a transgenic dwarf plant said method comprising transforming a plant with the antisense or sense construct of a *GA4H* gene or cDNA.

37. A method of making GA4H protein wherein said GA4H protein is substantially free of other *A. thaliana* proteins, comprising:

5

- a) transforming a prokaryotic or eukaryotic cell with a GA4H recombinant expression vector encoding a GA4H protein,
- b) expressing said GA4H protein, and
- c) purifying said GA4H protein substantially free of other *A. thaliana* proteins.

10

38. An antibody or fragment thereof, capable of binding a GA4H protein.

39. A method of identifying cells or tissues expressing GA4H comprising the steps of:

15

- a) incubating said cells or said tissues with an agent capable of binding to the GA4H protein or the RNA encoding GA4H; and
- b) detecting the presence of the bound agent.

40. The method of claim 39 wherein said agent capable of binding to the GA4H protein is an antibody or fragment thereof.

1/19

1	ATAAGAAAAAACACAAACATCTATCAAATTTACAAAGTTTTAAACTAATTAATAAG	60
61	AGCAAGATGCCCTGCTATGTAAACAGATGTGTTAGAGGCCATCCCATTACCTCCCACAC	120
1	M P A M L T D V F R G H P I H L P H	18
121	TCTCACATACCTGACTTCACATCTCTCCGGGAGCTCCCGGATTCTTACAAGTGACCCCT	180
19	S H I P D F T S L R E L P D S Y K W T P	38
181	AAAGACGATCTCCTCTTCTCCGCTGCTCCTTCTCCTCCGGCCACCGGTGAAAACATCCCT	240
39	K D D L L F S A A P S P P A T G E N I P	58
241	CTCATCGACCTCGACCACCCGGACGGACTAACCAAATCCGTCATGCATGTAGAACTTGG	300
59	L I D L D H P D A T N Q I G H A C R T W	78
301	GGTGCCTTCCAAATCTCAAACCACGGCGTCCCTTTGGGACTTCTCCAAGACATTGAGTTT	360
79	G A F Q I S N H G V P L G L L Q D I E F	98
361	CTCACC GG TAGTCTCTTCCGGCTACCTGTCCAACGCAAGCTTAAGTCTGCTCGGTCCGAG	420
99	L T G S L F G L P V Q R K L K S A R S E	118
421	ACAGGTGTGTCGGGCTACGGCGTGGCTCGTATCGCATCTTTCTTCAATAAGCAAATGTGG	480
119	T G V S G Y G V A R I A S F F N K Q M W	138
481	TCCGAAGGTTTCACCATCACTGGCTCGCCTCTCAACGATTTCCGTAACTTTGGCCCCAA	540
139	S E G F T I T G S P L N D F R K L W P Q	158
541	CATCACCTCAACTACTGCGATATCGTTGAAGAGTACGAGGAACATATGAAAAAGTTGGCA	600
159	H H L N Y C D I V E E Y E E H M K K L A	178
601	TCGAAATTGATGTGGTTAGCACTAAATTCAGTTGGGGTCAGCGAAGAAGACATTGAATCG	660
179	S K L M W L A L N S L G V S E E D I E W	198
661	GCCAGTCTCAGTTCAGATTTAACTGGGCCAAGCTGCTCTCCAGCTAAATCACTACCCG	720
199	A S L S S D L N W A Q A A L Q L N H Y P	218
721	GTTTGTCTGAACCGGACCGAGCCATGGGTCTAGCAGCTCATACCGACTCCACCCTCCTA	780
219	V C P E P D R A M G L A A H T D S T L L	238
781	ACCATTCTGTACCAGAACAAATACCGCCGGTCTACAAGTATTTCCGGATGATCTTGGTTGG	840
239	T I L Y Q N N T A G L Q V F R D D L G W	258
841	GTCACCGTGCCACCGTTTCTGGCTCGCTCGTGGTTAACGTTGGTGACCTCTTCCACATC	900
259	V T V P P F P G S L V V N V G D L F H I	278
901	CTATCCAATGGATTGTTTAAAGCGTGTTCACCGCGCTCCGGTTAACCAAACCAGAGCC	960
279	L S N G L F K S V L H R A R V N Q T R A	298

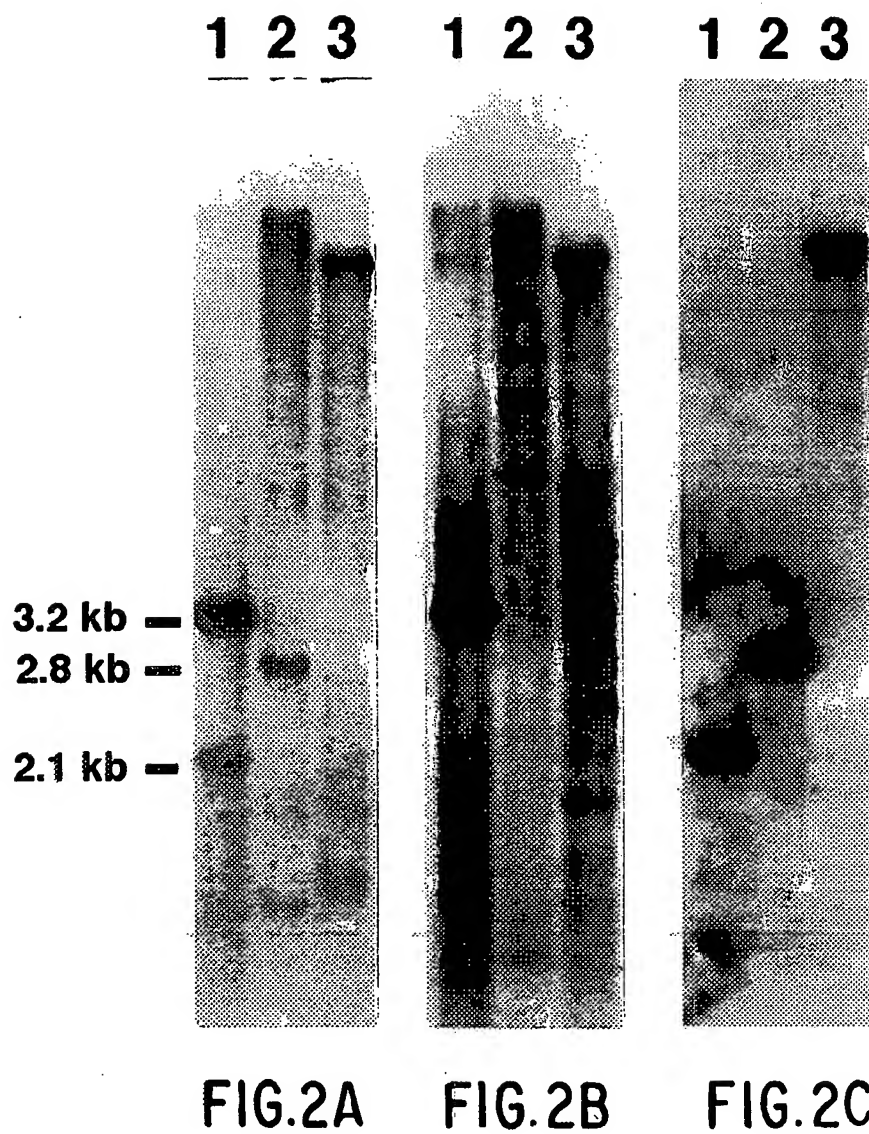
FIG. 1A

2/19

961	CGGTTATCTGTAGCATTCCCTTTGGGGTCCGCAATCTGATATCAAGATATCACCTGTACCG	1020
299	R L S V A F L W G P Q S D I K I S P V P	318
1021	AAGCTGGTTAGTCCCGTTGAATCGCCTCTATACCAATCGGTGACATGGAAAGAGTATCTT	1080
319	K L V S P V E S P L Y Q S V T W K E Y L	338
1081	CGAACAAAAGCAACTCACTTCAACAAAGCTCTTTCAATGATTAGAAATCACAGAGAAGAA	1140
339	R T K A T H F N K A L S M I R N H R E E	358
1141	TGATTAGATAATAATAGTTGTGATCTACTAGTTAGTTTGATTAAATAAATTGTTGTAAATG	1200
1201	ATTCAGCAATATGATTGTTTGTCTC	1228

FIG.1B

3/19



4/19

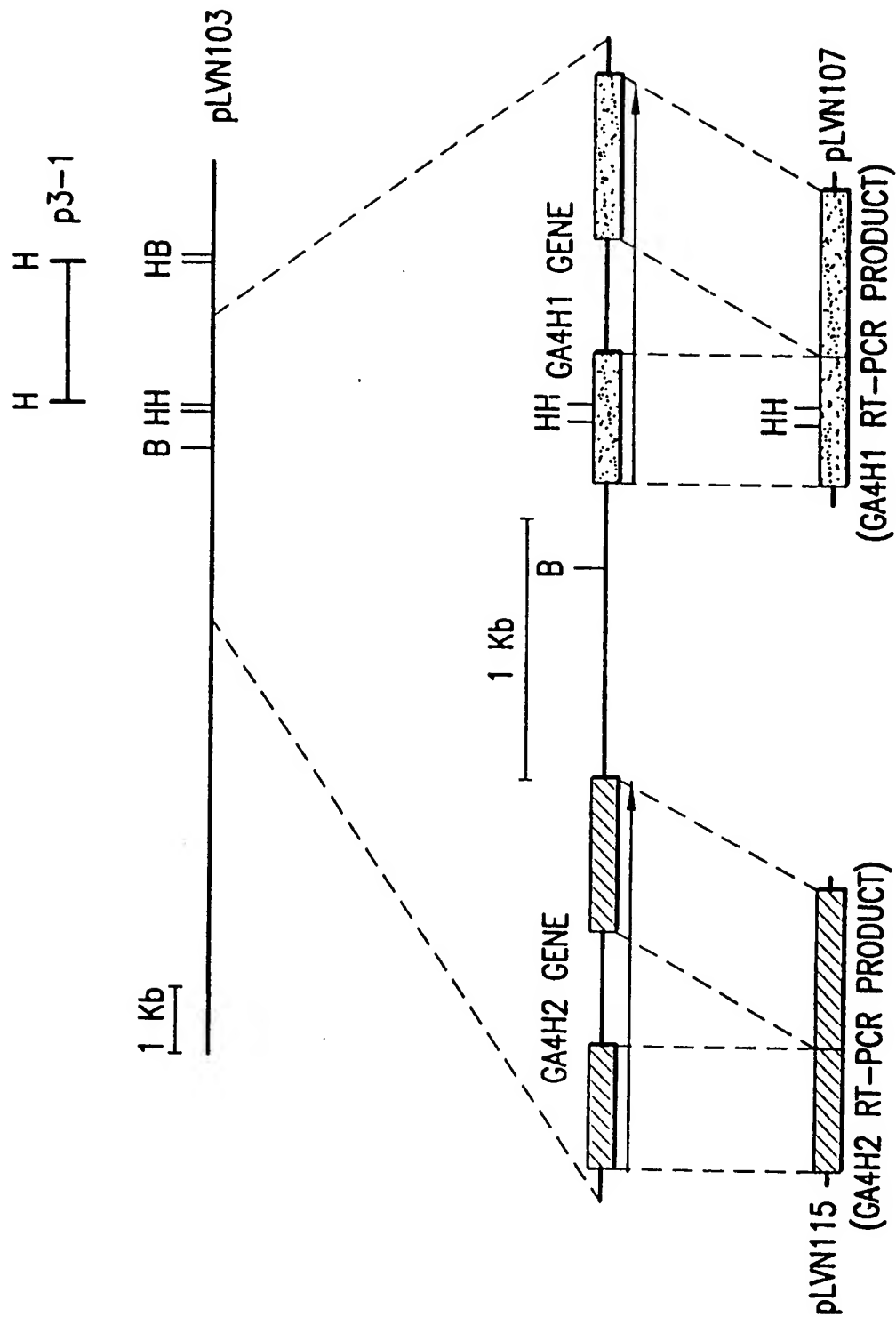


FIG.3

5/19

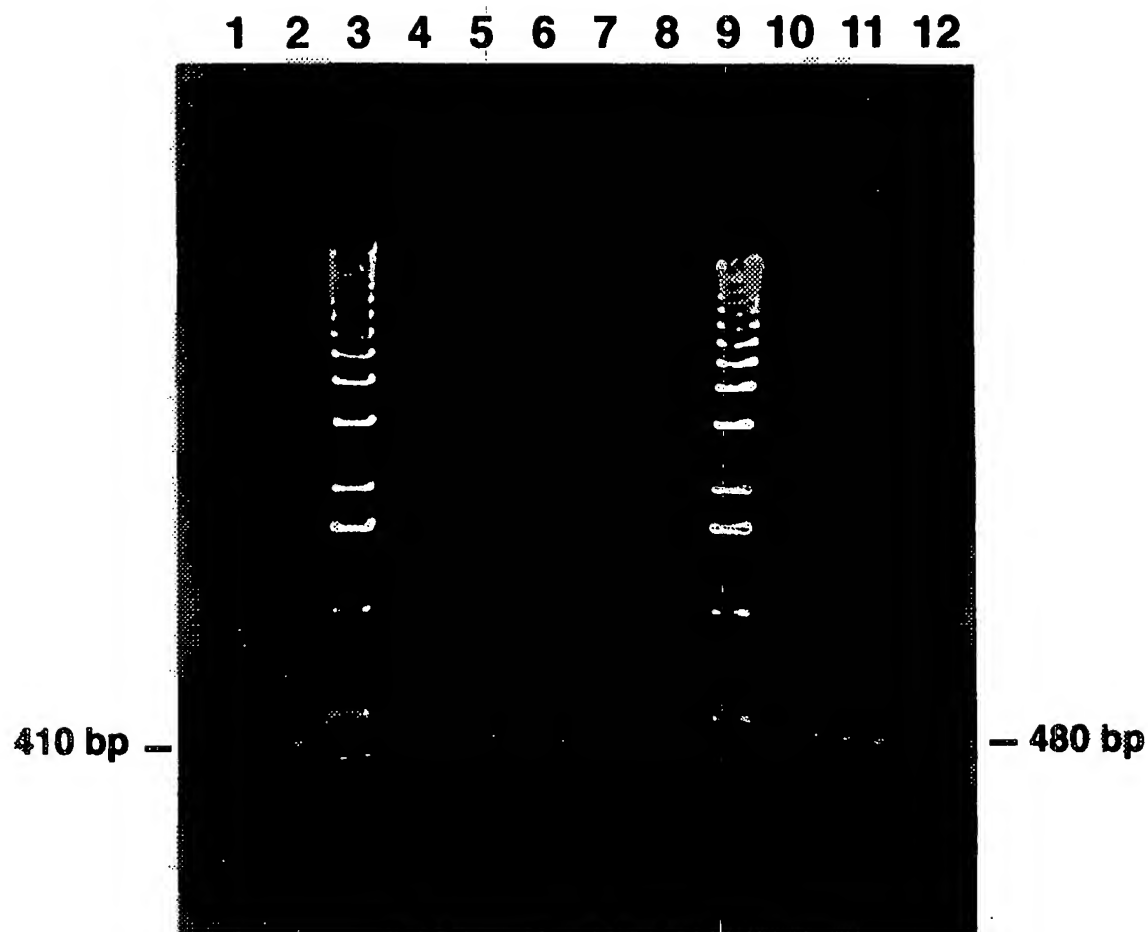


FIG.4A



6/19

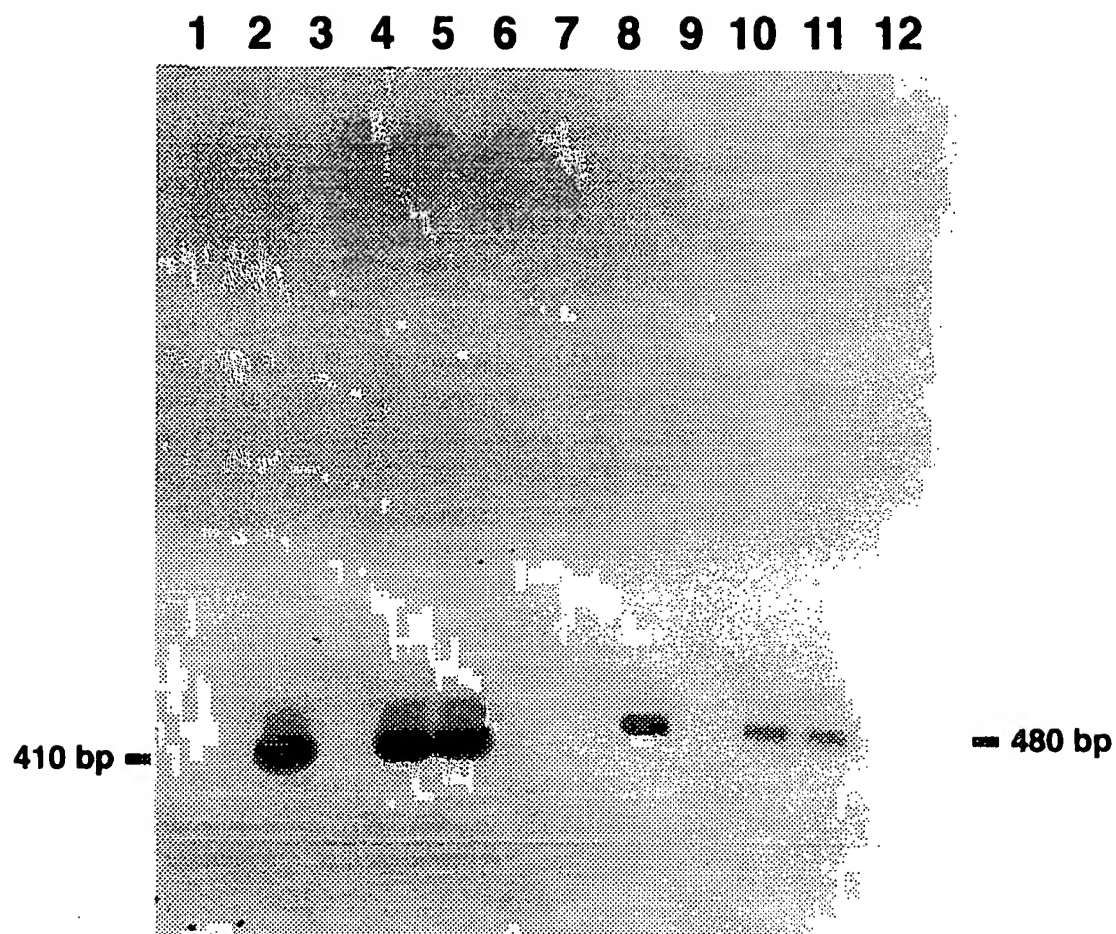


FIG.4B

7/19

1 glllalglo lgagcalccc olclclcal logllcacoo glcATGCCTT  
51 CACTAGCAGA AGAGATATGT ATTGGTAACT TAGGCAGTCT CCAAACACTC  
101 CCCGAGTCGT TCACCTGGAA ACTCACAGCC GCCGACTCCC TTCTGCGTCC  
151 CTCCTCCGCC GTCTATTTCG ACGCAGTGGG AGAGTCCATT CCTGTGATCG  
201 ACCTCTCTAA TCCTGACGTT ACCACCCTCA TTGGAGATGC CTCCAAAACA  
251 TGGGGAGCGT TTCAGATAGC CAACCACGGG ATTTCTCAGA AGCTTCTCGA  
301 TGATATCGAG TCTCTGTCCA AAACCTATT CGACATGCCG TCAGAGAGGA  
351 AGCTTGAAGC GGCTTCCTCC GATAAAGGAG TTAGTGGCTA CGGAGAACCT  
401 CGAATCTCCC CCTTTTTCGA GAAGAAAATG TGGTCTGAAG GGTTTACTAT  
451 TGCCGATGAC TCCTACCGCA ACCATTTCAG TACTCTTTGG CCTCATGATC  
501 ACACCAAGTA CTGCGGTATA ATCCAAGAAT ACGTGGACGA AATGGAAAAA  
551 TTAGCAAGCA GACTTCTGTA TTGCACATTA GGCTCACTTG GTGTCACCGT  
601 GGAAGACATT GAATGGGCTC ACAAGCTAGA GAAATCTGGA TCAAAAGTGG  
651 GCAGAGGGCC CATAAGACTA AACCCTACC CGGTTTGTCC TGAACCAGAA  
701 CGAGCCATGG GTCTAGCCGC TCATACAGAC TCCACTATCC TAACCATTC  
751 GCACCAGAGC AACACGGGAG GGCTACAAGT GTTCAGGGAA GAGTCCGGTT  
801 GGGTCACCGT TGAGCCGGCT CCTGGTGTCC TCGTGGTCAA CATGGGTGAT  
851 CTCTTTCACA TCTTATCGAA CGGAAAATC CCAAGCGTGG TTCATCGAGC  
901 CAAAGTTAAC CATACTCGGT CAAGAATTTC GATTGCGTAC TTATGGGGTG  
951 GTCCAGCTGG TGATGTGCAA ATCGCACCTA TCTCTAAGTT AACCGGTCCG  
1001 GCTGAACCGT CTCTTTACCG GTCAATTACA TGGAAAGAGT ATCTCCAAAT  
1051 AAAGTATGGG GTTTTCGACA AGGCCATGGA CGCAATTAGG GTCGTTAATC  
1101 CCACCAATga galclclcllc lcaoolclcl lcl

FIG.5

	GA-P15	GA-P14S	
acolatgltgltgltatclatgcolatolatccoaogtaallglllatgltgatgogcalc			60
ccollclclcatllogltlcacaaoglcATGCCTTCACTAGCAGAAGAGATATGTATTGGTAA			120
	M P S L A E E I C I G N		12
CTTAGGCAGTCTCCAAACACTCCCCGAGTCGTTACCTGGAAACTCACAGCCGCCGACTC			180
L G S L Q T L P E S F T W K L T A A D S			32
CCTTCTGCGTCCCTCCTCCGCCGTCTCATTGACGCAGTGAAGAGTCCATTCCCTGTGAT			240
L L R P S S A V S F D A V E E S I P V I			52
CGACCTCTCTAATCCTGACGTTACCACCCTCATTGGAGATGCCTCCAAAACATGGGGAGC			300
D L S N P D V T T L I G D A S K T W G A			72
GTTTCAGATAGCCAACCACGGGATTTCTCAGAAGCTTCTCGATGATATCGAGTCTCTGTC			360
F Q I A N H G I S Q K L L D D I E S L S			92
CAAAACCCTATTGACATGCCGTCAGAGAGGAAGCTTGAAGCGGCTTCCTCCGATAAAGG			420
K T L F D M P S E R K L E A A S S D K G			112
AGTTAGTGGCTACGGAGAACCTCGAATCTCCCCCTTTTTCGAGAAGAAAATGTGGTCTGA			480
V S G Y G E P R I S P F F E K K M W S E			132
	GA-P2	GA-P13	
AGGGTTTACTATTGCCGATGACTCCTACCGCAACCATTCAATACTCTTTGGCCTCATGA			540
G F T I A D D S Y R N H F N T L W P H D			152
TCACACCAAGTACTGgtaacgtclatlacacacacalatalatalallllllgcllalllclg			600
H T K Y W			157
caaaogltgtggcaaoggaallgcacaccllllllllllgcaclaogocllaglltallallo			660
aaogltglltaaatgllllllllclgllicalaanaaogltlllatatgllccgagtaallga			720
tglllatgallaglgalaoclgalaocacalogaglglogccllcaaoglllcltaallao			780
ataglltgagcaacalccllatalalllallgoogtaglactllcllalllgcolalllocagca			840
oalltaaoglaaccaogltclclalgaaalglgalaalllggcltaalgclgaggglclltaaco			900
llagallaccaaaaaocclltaallacglglaaallglallllgclllllclalllllgglallgt			960
		GA-P6	
gcagCGGTATAATCCAAGAATACGTGGACGAAATGGAAAAATTAGCAAGCAGACTTCTGT			1020
G I I Q E Y V D E M E K L A S R L L Y			176

FIG. 6A

ATTGCACATTAGGCTCACTTGGTGTACCCGTGGAAGACATTGAATGGGCTCACAAAGCTAG 1080  
C T L G S L G V T V E D I E W A H K L E 196

AGAAATCTGGATCAAAAGTGGGCAGAGGCGCCATACGACTAAACCACTACCCGGTTTGTCT 1140  
K S G S K V G R G A I R L N H Y P V C P 216  
GA-P17

CTGAACCAGAACGAGCCATGGGTCTAGCCGCTCATACAGACTCCACTATCCTAACCATTCT 1200  
E P E R A M G L A A H T D S T I L T I L 236

TGCACCAGAGCAACACCGGAGGGCTACAAGTGTTTCAGGGAAGAGTCCGGTTGGGTCACGG 1260  
H Q S N T G G L Q V F R E E S G W V T V 256

TTGAGCCCGGCTCCTGGTGTCTCGTGGTCAACATGGGTGATCTCTTTCACATCTTATCGA 1320  
E P A P G V L V V N M G D L F H I L S N 276

ACGGGAAAATCCCAAGCGTGGTTCATCGAGCCAAAGTTAACCATACTCCGTCAAGAATTT 1380  
G K I P S V V H R A K V N H T R S R I S 296

CGATTGCGTACTTATGGGGTGGTCCAGCTGGTGATGTGCAAATCGCACCTATCTCTAAGT 1440  
I A Y L W G G P A G D V Q I A P I S K L 316

TAAACCGGTCCGGCTGAACCGTCTCTTTACCGGTCAATTACATGGAAAGAGTATCTCCAAA 1500  
T G P A E P S L Y R S I T W K E Y L Q I 336

TAAAGTATGAGGTTTTCGACAAGGCCATGGACGCAATTAGGGTCGTTAATCCCACCAATL 1560  
K Y E V F D K A M D A I R V V N P T N 355

aaatctccttctcaaatactctcttaaatgaaaaaacctaaatttaaatgcga 1610  
GA-P1X GA-P16

FIG. 6B

10/19

1 lcalaccaaa aacalaagc caaaalalaa acacalaagc clilllagcAI  
51 GAGTTCAACG TTGAGCGATG TGTITAGATC GCATCCCATT CACATCCCAC  
101 TCTCAAACCC ACCTGACTTC AAATCTCTCC CGGATTCTTA CACGTGGACT  
151 CCTAAAGATG ATCTCCTCTT CTCCGCCCTCC GCCTCCGACG AAACCCTGCC  
201 GCTCATCGAC CTCTCCGATA TCCACGTGGC CACTCTTG TG GGCCATGCTT  
251 GTACCACGTG GGGAGCGTTC CAGATCACCA ACCACGGCGT CCCCTCCGA  
301 CTCTCGACG ACATTGAGTT CCTCACC GGA AGTCTTTTCC GGCTTCCGT  
351 ACAGCGGAAG CTCAAGGCGG CTCCGTCAGA GAATGGCGTC TCCGGCTACG  
401 GCGTAGCTCG TATTGCTTCG TTCTTTAATA AGAAGATGTG GTCCGAAGGT  
451 TTCACCGTTA TTGGCTCTCC CCTCCACGAT TTCCGTAAAC TCTGGCCCAG  
501 CCACCACCTC AAATACTGTG AAATTATTGA AGAGTATGAA GAACATATGC  
551 AAAAGTTGGC AGCCAAGTTG ATGTGGTTCC CATTAGGTTT ACTGGGAGTT  
601 GAAGAAAAGG ACATACAATG GGCCGGGCCT AATTCAGACT TTCAAGGAAC  
651 CCAAGCAGCT ATCCAAC TAA ACCATTATCC AAAATGTCCA GAACCAGACA  
701 GAGCCATGGG CCTCGCAGCC CATA CAGACT CGACCCTCAT GACCATTCTG  
751 TACCAGAACA ACACCGCCGG TCTCCAAGTT TTCCGGGATG ACGTGGGCTG  
801 GGTACCGCG CCACCTGTCC CTGGCTCGCT GGTGGTCAAC GTCGGTGACT  
851 TGCTCCACAT TTTAACCAAC GGAATCTTCC CGAGCGTGCT TCACCGAGCC  
901 AGGGTTAACC ACGTCCGATC TCGGTCTCA ATGGCTTACC TGTGGGGTCC  
951 ACCATCCGAT GTAATGATCT CTCCACTTCC CAACTGGTT GATCCTCTCC  
1001 AATCTCTCT CTACCCATCT CTCACTTGGG AACAATACCT TGCTACCAAA  
1051 GCTACTCATT TTAATCAATC TCTTTCCATT ATTAGAAATI aaclglcllc  
1101 cgact

FIG.7

11/19

		GA-P27E	
lcaccgctclalaaatcacclccclcllclccaccocaaaglalcolatcalaccocaaacac	60		
laaogccaaalalaaacacaloagccllllagcATGAGTTCAACGTTGACCGATGTGTT	120		
M S S T L S D V F	9		
TAGATCGCATCCCATTCACATCCCACTCTCAAACCCACCTGACTTCAAATCTCTCCCGGA	180		
R S H P I H I P L S N P P D F K S L P D	29		
TTCTTACACGTGGACTCCTAAAGATGATCTCTTCTCCGCCTCCGCCTCCGACGAAAC	240		
S Y T W T P K D D L L F S A S A S D E T	49		
CCTGCCGCTCATCGACCTCTCCGATATCCAGGTGGCCACTCTTGTTGGCCATGCTTGAC	300		
L P L I D L S D I H V A T L V G H A C T	69		
CACGTGGGGAGCGTTCCAGATCACCAACCACGGCGTCCCCTCGCGACTTCTCGACGACAT	360		
T W G A F Q I T N H G V P S R L L D D I	89		
		GA-P20	
TGAGTTCCTACCGGAAGTCTTTTCCGGCTTCCCGTACAGCGGAAGCTCAAGCGGCTCG	420		
E F L T G S L F R L P V Q R K L K A A R	109		
GTCAGAGAATGGCGTCTCCGGCTACGGCGTAGCTCGTATTGCTTCGTTCTTTAATAAGAA	480		
S E N G V S G Y G V A R I A S F F N K K	129		
GATGTGGTCCGAAGGTTTCACCGTTATTGGCTCTCCCCTCCACGATTTCGGTAAACTCTG	540		
M W S E G F T V I G S P L H D F R K L W	149		
GCCCAGCCACCACCTCAAATACTGgtatclllllcaalggllcallllclcoacgllloag	600		
P S H H L K Y W	157		
accatoltaacglaacglaaccllclllglolgaaaaaaaaaaaaaaaactglggacgl	660		
loglocagllgacclatlcoallgotalogolcgggaoloolacgaaaoggglaaoglog	720		
aaaccallllllgccolglcglagllagloaaogcacoalgaooacclcalggaccacc	780		
aaaaogallacalgotatolalolalolalolalllclololaaalallolaloolalall	840		
		GA-P19	
lolaloolalllaltglcooooallaaolgoaaalaaolalllalcaggogaaltgtgaalo	900		
caglotaaogallllclllggclacalgaogalllclalagalllgaaggltaagolact	960		
oalllcalalllclgallicoacogTGAATTATTGAAGAGTATGAAGAACATATGCAAAA	1020		
E I I E E Y E E H M Q K	169		
GTTGGCAGCCAAGTTGATGTGCTTCGATTAGGTTCoCTGGGAGTTGAAGAAAAGGACAT	1080		
L A A K L M W F A L G S L G V E E K D I	189		

FIG.8A

SI RSTITI I TE SHEET (RI I I E 26)

12/19

ACAATGGGCGGGCCTAATTCAGACTTTCAAGGAACCCAAGCAGCTATCCAATAACCA	1140
Q W A G P N S D F Q G T Q A A I Q L N H	209
TTATCCAAATGTCCAGAACCAGACAGAGCCATGGGCCTCGCAGCCCATACAGACTCGAC	1200
Y P K C P E P D R A M G L A A H T D S T	229
CCTCATGACCATTCTGTACCAGAACAACACCGCCGGTCTCCAAGTTTCCGGGATGACGT	1260
L M T I L Y Q N N T A G L Q V F R D D V	249
GGGCTGGGTACCGCGCCACCTGTCCCTGGCTCGCTGGTGGTCAACGTCCGTGACTTGCT	1320
G W V T A P P V P G S L V V N V G D L L	269
CCACATTTTAACCAACGGAATCTTCCCGAGCGTGCTTCACCGAGCCAGGGTTAACCACGT	1380
H I L T N G I F P S V L H R A R V N H V	289
CCGATCTCGGTTCTCAATGGCTTACCTGTGGGTCCACCATCCGATGTAATGATCTCTCC	1440
R S R F S M A Y L W G P P S D V M I S P	309
ACTTCCCAAACCTGGTTGATCCTCTCCAATCTCTCTACCCATCTCTCACTTGGAAACA	1500
L P K L V D P L Q S P L Y P S L T W K Q	329
ATACCTTGCTACCAAAGCTACTCATTTTAATCAATCTCTTTCCATTATTAGAAATlaocl	1560
Y L A T K A T H F N Q S L S I I R N	347
glcttccgacigaatttcttgollltcagollltactlalltlltctllagtaatatgol	1620
← GA-P21	
gatatctalltactglltcgollltagolgggtggttcllcaaalccacoattagtagcll	1680
oolottgoll	1690

FIG.8B

13/19

Ga4	MPAMLTDFVR	GHPIHLPHSH	IPDFTSLREL	PDSYKW..TP	38
Ga4h2	MSSTLSDFVR	SHPIHIPLSN	PPDFKS...L	PDSYTW..TP	35
Ga4h1	MPSLAEEIC.	.....	IGNLGSLQTL	PESFTWKLTA	29
Ga4	KDDLIF.SAA	PSPPATGENT	PLIDLHPDA	TNQIGHACRT	77
Ga4h2	KDDLIF.SAS	AS....DETL	PLIDLSDIHV	ATLVGHACTT	70
Ga4h1	ADSLLRPSSA	VSFDAVEESI	PVIDLSNPDV	TTLIGDASKT	69
Ga4	WGAFAQISNHG	VPLGLLQDIE	FLTGSLFGLP	VQRKLKSARS	117
Ga4h2	WGAFAQITNHG	VPSRLLDDIE	FLTGSLFRLP	VQRKLKAARS	110
Ga4h1	WGAFAQIANHG	ISQKLLDDIE	SLSKTLFDMP	SERKLEAASS	109
Ga4	ETGVSGYGVA	RIASFFNKQM	WSEGFTITGS	PL.NDFRKLW	156
Ga4h2	ENGVSFGYVA	RIASFFNKQM	WSEGFTVIGS	PL.HDFRKLW	149
Ga4h1	DKGVSGYGEP	RISPFFEKKM	WSEGFTIADD	SYRNHFNTLW	149
Ga4	PQHHLNYCDI	VEEYEEHMKK	LASKLMWLAL	NSLGVSEEDI	196
Ga4h2	PSHHLKYCEI	IEEYEEHMOK	LAACKLMWFAL	GSLGVEEKDI	189
Ga4h1	PHDHTKYCGI	IQEYVDEMEK	LASRLLYCTL	GSLGVTVEDI	189
Ga4	EWAS..LSSD	LNWAQAAQL	NHYPVCPEPD	RAMGLAAHTD	234
Ga4h2	QWAG..PNSD	FQGTQAAIQL	NHYPKCPEPD	RAMGLAAHTD	227
Ga4h1	EWAKLEKSG	SKVGRGKIRL	NHYPVCPEPE	RAMGLAAHTD	229
Ga4	STLLTILYQN	NTAGLQVFRD	DLGWVTVPF	PGSLVVNVGD	274
Ga4h2	STLMTILYQN	NTAGLQVFRD	DVGWVTAPPV	PGSLVVNVGD	267
Ga4h1	STILTILHQS	NTGGLOVFRE	ESGWVTVEPA	PGVLVVNMGD	269
Ga4	LFHILSNGLF	KSVLHRARVN	QTRARLSVAF	LW.GPQSDIK	313
Ga4h2	LLHILTNGIF	PSVLHRARVN	HVRSRFSMAY	LW.GPPSDVM	306
Ga4h1	LFHILSNGKI	PSVVHRAKVN	HTRSRIISIAV	LWGGPAGDVQ	309
Ga4	ISPVPKLVSP	VESPLYQSVT	WKEYLRITKAT	HFNKALSMIR	353
Ga4h2	ISPLPKLVDP	LQSPLYPSLT	WKQYLATKAT	HFNQSLSIIR	346
Ga4h1	IAPISKLTGP	AEPSLYRSIT	WKEYLQIKYG	VFDKAMDAIR	349
Ga4	NHREE.	358			
Ga4h2	N.....	347			
Ga4h1	VVNPTN	355			

FIG.9  
SUBSTITUTE SHEET (RULE 26)



14/19

GA4H2 (ARAB.)	GA4	GA4H2	GA4H1	GA5	F3H	FLS	ANS
	76 (85)	57 (73)	56 (73)	29 (49)	30 (49)	31 (52)	29 (52)
GA4H1 (ARAB.)				30 (54)	29 (50)	29 (51)	29 (50)
				30 (56)	27 (52)	31(57)	30 (55)
				32 (57)	31 (56)	29 (55)	32 (54)
				29 (60)	30 (59)	25(52)	29 (54)
							31 (55)
							37 (64)
							32 (57)

FIG.10

15/19

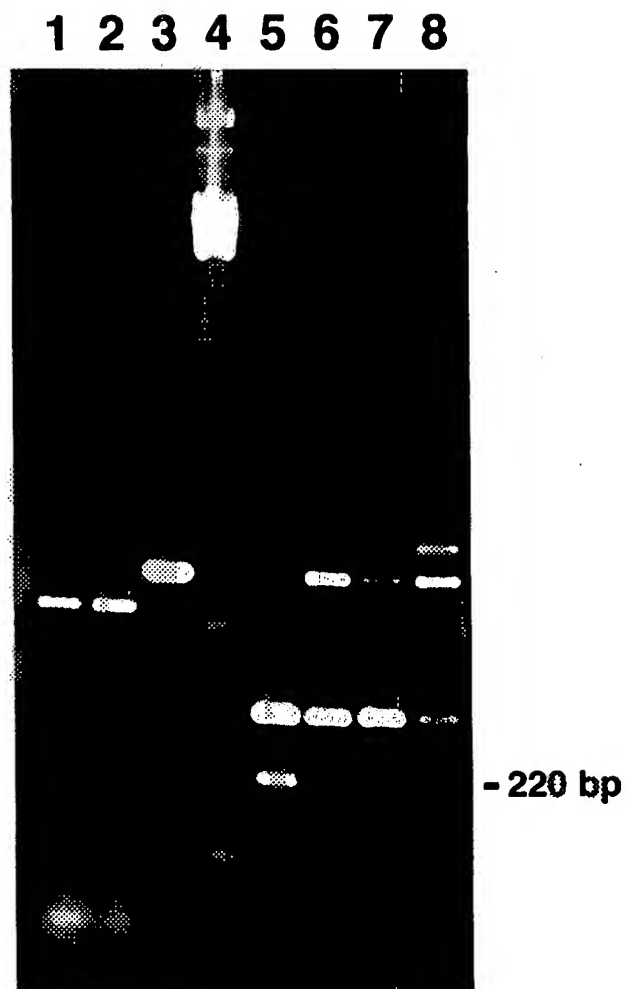


FIG.11A

16/19

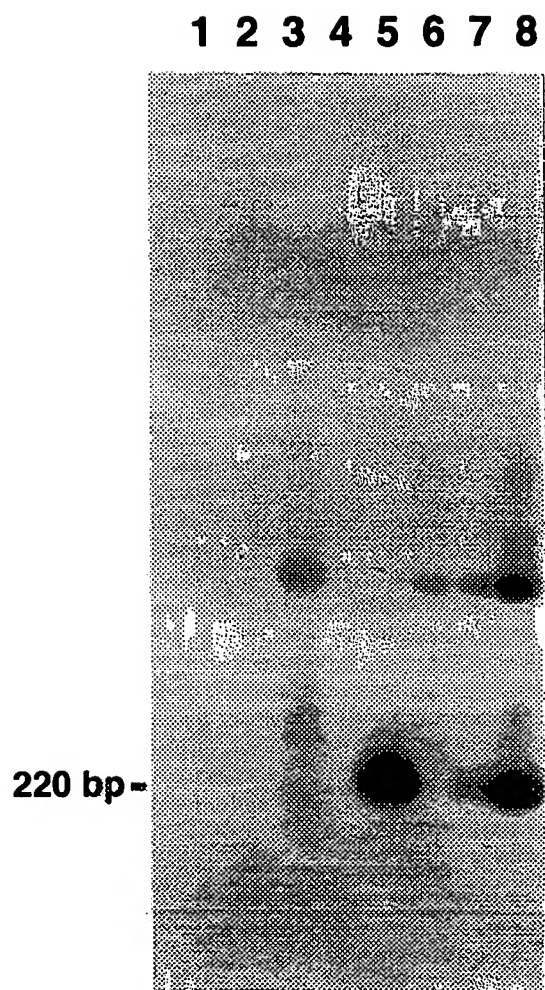


FIG.11B

17/19

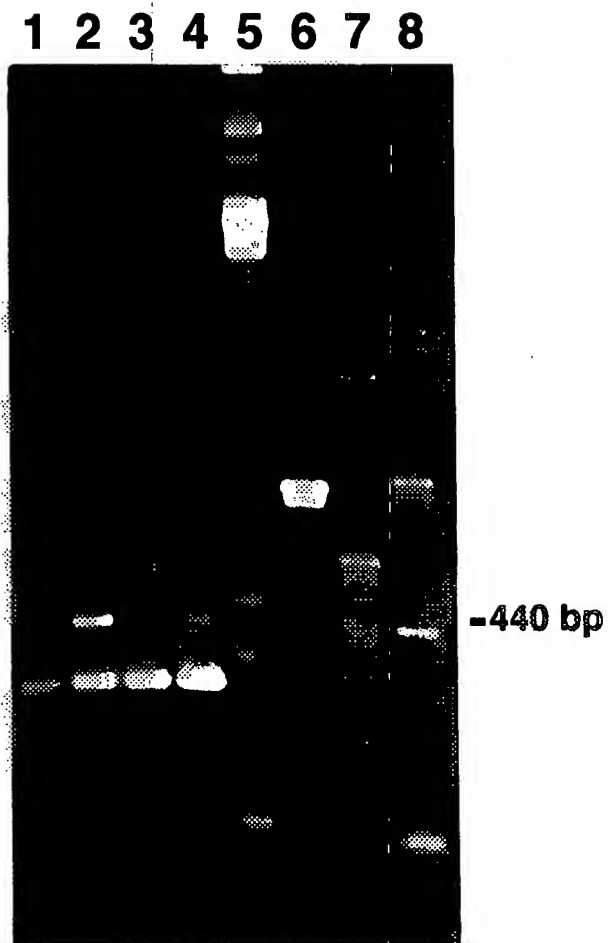


FIG.12A

18/19

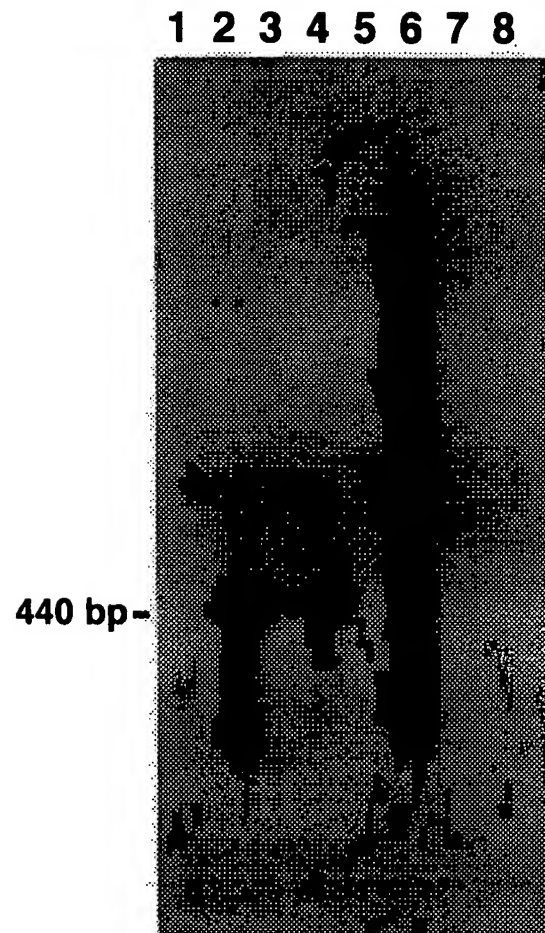
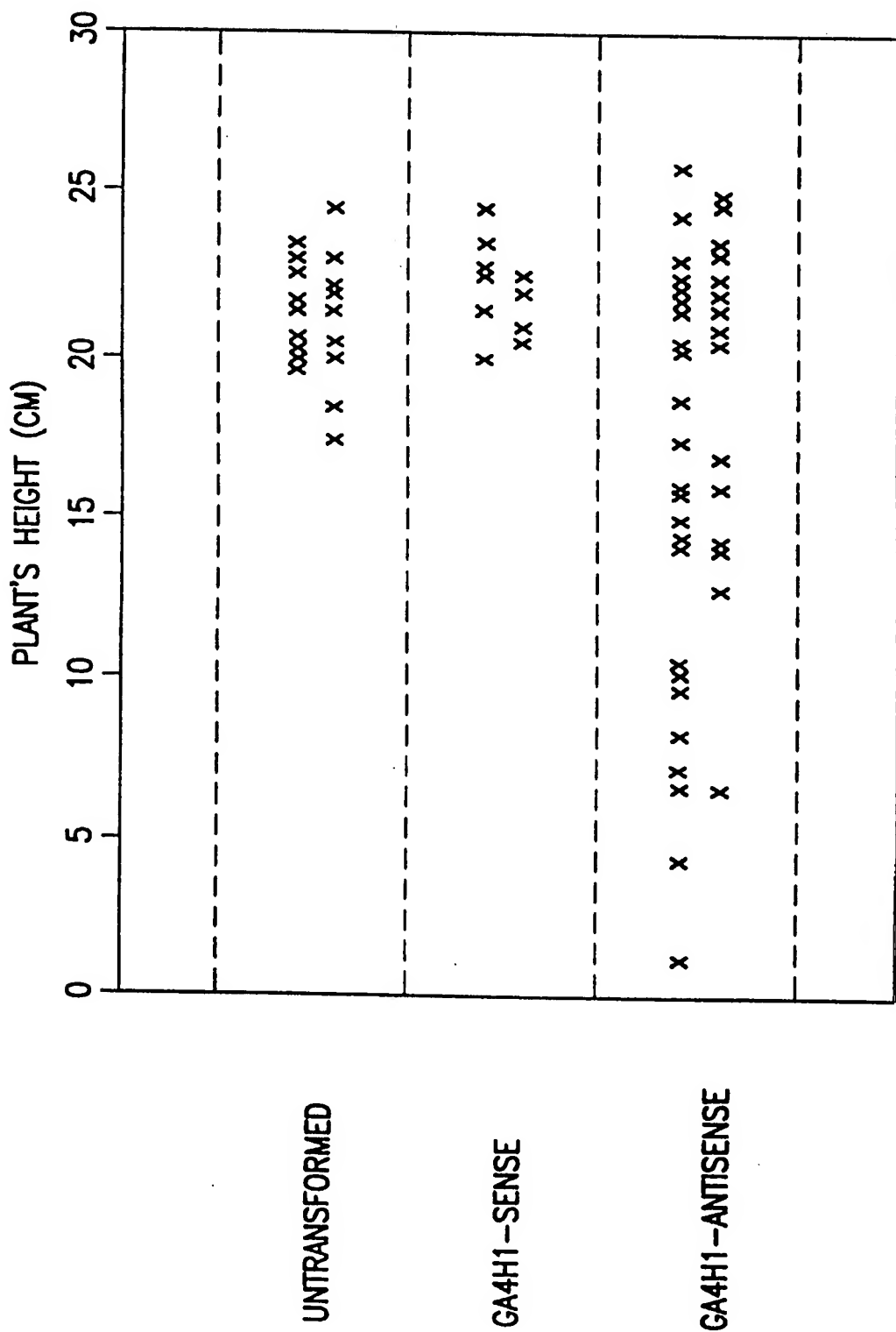


FIG.12B

19/19



x ONE PLANT

FIG.13

Applicant's or agent's file reference number	0609.439PC01	International application No.	TBA
---	--------------	-------------------------------	-----

**PCT/US 98/13044**

**INDICATIONS RELATING TO A DEPOSITED MICROORGANISM**  
(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>4</u> , line <u>25-26</u> .	
<b>B. IDENTIFICATION OF DEPOSIT</b>	
Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution	
American Type Culture Collection	
Address of depositary institution (including postal code and country)	
12301 Parklawn Drive Rockville, Maryland 20852 United States of America	Now at 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America
Date of deposit May 20, 1997	Accession Number ATCC 98436
<b>C. ADDITIONAL INDICATIONS</b> (leave blank if not applicable)	
This information is continued on an additional sheet <input type="checkbox"/>	
Arabidopsis thaliana genomic DNA of GA4H1 and GA4H2 genes cloned into pBSKS(+) (Stratagene) vector pLVN103 in DH5α	
In respect of those designations in which a European Patent is sought a sample of the deposited microorganism will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample (Rule 28(4) EPC).	
<b>D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE</b> (if the indications are not for all designated States)	
<b>E. SEPARATE FURNISHING OF INDICATIONS</b> (leave blank if not applicable)	
The indications listed below will be submitted to the international Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

<b>For receiving Office use only</b>	<b>For International Bureau use only</b>
<input type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

*(Arabidopsis thaliana genomic DNA of GA4H1 and  
GA4H2 genes cloned into pBSKS(+)  
(Stratagene) vector pLVN103 in DH5 $\alpha$ )*

**NETHERLANDS**

The applicant hereby requests that until the date of a grant of a Netherlands patent or until the date on which the application is refused or withdrawn or lapsed, the microorganism shall be made available as provided in Rule 31F(1) of the Patent Rules only by the issue of a sample to an expert. The request to this effect must be furnished by the applicant with the Netherlands Industrial Property Office before the date on which the application is made available to the public under Section 22C or Section 25 of the Patents Act of the Kingdom of the Netherlands, whichever of the two dates occurs earlier.

**NORWAY**

The applicant hereby requests that, until the application has been laid open to public inspection (by the Norwegian Patent Office), or has been finally decided upon by the Norwegian Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the Norwegian Patent Office not later than at the time when the application is made available to the public under Sections 22 and 33(3) of the Norwegian Patents Act. If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Norwegian Patent office or any person approved by the applicant in the individual case.

**SINGAPORE**

The applicant hereby requests that the furnishing of a sample of a microorganism shall only be made available to an expert. The request to this effect must be filed by the applicant with the International Bureau before the completion of the technical preparations for international publication of the application.

**SWEDEN**

The applicant hereby requests that, until the application has been laid open to public inspection (by the Swedish Patent Office), or has been finally decided upon by the Swedish Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the International Bureau before the expiration of 16 months from the priority date (preferably on the Form PCT/RO/134 reproduced in annex Z of Volume I of the PCT Applicant's Guide). If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Swedish Patent office or any person approved by the applicant in the individual case.

**UNITED KINGDOM**

The applicant hereby requests that the furnishing of a sample of a microorganism shall only be made available to an expert. The request to this effect must be filed by the applicant with the International Bureau before the completion of the technical preparations for international publication of the application.



*(Arabidopsis thaliana genomic DNA of GA4H1 and  
GA4H2 genes cloned into pBSKS(+)  
(Stratagene) vector pLVN103 in DH5 $\alpha$ )*

#### AUSTRALIA

The applicant hereby gives notice that the furnishing of a sample of a microorganism shall only be effected prior to the grant of a patent, or prior to the lapsing, refusal or withdrawal of the application, to a person who is a skilled addressee without an interest in the invention (Regulation 3.25(3) of the Australian Patents Regulations).

#### CANADA

The applicant hereby requests that, until either a Canadian patent has been issued on the basis of the application or the application has been refused, or is abandoned and no longer subject to reinstatement, or is withdrawn, the furnishing of a sample of deposited biological material referred to in the application only be effected to an independent expert nominated by the Commissioner of Patents.

#### DENMARK

The applicant hereby requests that, until the application has been laid open to public inspection (by the Danish Patent Office), or has been finally decided upon by the Danish Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the Danish Patent Office not later than at the time when the application is made available to the public under Sections 22 and 33(3) of the Danish Patents Act. If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Danish Patent office or any person approved by the applicant in the individual case.

#### FINLAND

The applicant hereby requests that, until the application has been laid open to public inspection (by the National Board of Patents and Registration), or has been finally decided upon by the National Board of Patents and Registration without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the International Bureau before the expiration of 16 months from the priority date (preferably on the Form PCT/RO/134 reproduced in annex Z of Volume I of the PCT Applicant's Guide). If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the National Board of Patents and Registration or any person approved by the applicant in the individual case.

#### ICELAND

The applicant hereby requests that, until the application has been laid open to public inspection (by the Icelandic Patent Office), or has been finally decided upon by the Icelandic Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art.

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC 6 C12N15/53 C12N9/02 C12N5/10 C12N15/11 C07K16/40  
 G01N33/563 C12N15/82 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K C12Q G01N A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 96 05317 A (GEN HOSPITAL CORP) 22 February 1996 page 7,8; example 11; claims 18-21 ---	1-40
A	CHASAN, R.: "GA biosynthesis: a glimpse at the genes" THE PLANT CELL, vol. 7, February 1995, pages 141--143, XP002079878 see the whole document --- -/--	1-40

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

**\* Special categories of cited documents :**

"A" document defining the general state of the art which is not considered to be of particular relevance  
 "E" earlier document but published on or after the international filing date  
 "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)  
 "O" document referring to an oral disclosure, use, exhibition or other means  
 "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention  
 "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone  
 "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.  
 "&" document member of the same patent family

Date of the actual completion of the international search

15 October 1998

Date of mailing of the international search report

27/10/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
 Fax: (+31-70) 340-3016

Authorized officer

Holtorf, S

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>TALON, M., ET AL.: "endogenous gibberellins in Arabidopsis thaliana and possible steps blocked in the biosynthetic pathways of the semidwarf ga4 and ga5 mutants"</p> <p>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE USA, vol. 87, October 1990, pages 7983-7987, XP002079879 see figure 1</p> <p>-----</p>	1-40

# INTERNATIONAL SEARCH REPORT

international application No.

PCT/US 98/13044

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☒ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:  
see FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

### Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Claim 20 was read as referring to claim 19; claim 21 as referring to claim 20, and claim 22 as referring to claims 20-21.

Furthermore, the SEQIDs mentioned in the claims do not correspond to the SEQIDs of the filed sequence listings as far as the clone GA4H2 is concerned.

SEQIDs 8, 9, 10 in claims 3, 4, 17 and 21 should be read as SEQIDs 9, 10, 11, respectively.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9605317      A	22-02-1996	AU      3244995 A	07-03-1996
		CA      2197753 A	22-02-1996
		EP      0778895 A	18-06-1997
		JP      10508186 T	18-08-1998
<hr/>			